Multiple access stability and broadcast delay in wireless networks

A Thesis

Submitted to the Faculty

of

Drexel University

by

Nan Xie

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

August 2014

Dedications

To my parents: Heping Xie and Lijuan Jiang,

who have sacrificed so much for me

献给我的父母：谢和平 与 姜丽娟，

他们为我而牺牲了自己的太多

Acknowledgments

"Some people come into our lives and quickly go. Some people move our souls to dance. They awaken us to a new understanding with the passing whisper of their wisdom. Some people make the sky more beautiful to gaze upon. They stay in our lives for awhile, leave footprints on our hearts, and we are never, ever the same." – Flavia Weedn

My path to a Ph.D. is long and arduous. I would like to first and foremost thank my Ph.D. advisor Professor Steven Weber who has in many ways nurtured my intellectual growth. His quest for intuition behind problems and relentless pursuit of "the proof" not only improves the quality of my work in profound ways, it also serves as a constant source of inspiration. Furthermore, his insistence on and exemplification of writing well transforms the accessibility of our research work. Aside from my advisor, Professor John MacLaren Walsh has interacted with me perhaps more than any other faculty has. His imprints on me are irreplaceable, too. Over the years I have benefited to varying extents from my thesis committee members Professors Alejandro Ribeiro (University of Pennsylvania), Harish Sethu, Ali Shokoufandeh, and other faculty such as Professors Robert P. Boyer (Drexel University Department of Mathematics), Hugo J. Woerdeman (Drexel University Department of Mathematics), Athina Petropulu (Rutgers University), Saswati Sarkar (University of Pennsylvania), and Simon Foucart (University of Georgia).

Starting from graduate school, my Master's advisor Professor Mao Tian (Wuhan University) was the first person who guided me through research projects. Later on at the University of Florida, I was fortunate to have encountered Professors José A.B. Fortes, John M. Shea, Tan F. Wong, Ye Xia, and Shigang Chen. Of course, during my more than twenty-five years of formal education, many of the dedicated and inspiring teachers and mentors have touched my life, enlightening me and fostering my spirit of inquiry in far-reaching ways.

I am grateful to my former and current labmates at Drexel University who have made the environment of my work life stimulating and supportive: Ananth Kini, Ilaria Malanchini, Vijaya Pendyala, Zhen Zhao, Bradford Boyle, Alex Fridman, Congduan Li, Jeff Wildman, and Ni An.

In my personal life, I benefited from thought-provoking conversations with, among many others, Guannan Chen, Hengjun Cui, and Zongquan Gu. I am indebted to Dr. Jiangyuan Li, who is always ready for help and is my model of a true research scientist.

As an introverted person, I view the fifteen (or more) years' friendship with the following people as a blessing during my ups and downs: Jianlin Jiang, Caijun Wang, Yongjia Wang, Jie Xiang, and Dan Xu.

Last but not the least, I feel appreciative for the understanding and love from all my family members, in particular the unfailing love and support from my wife, Liyun Wu, as well as my parents-in-law, Linhua Wu and Mudi Xiao, despite the fact that they have been separated from me by the Pacific ocean over these many years. My achievements would not be possible without my family members. I never forget that my parents' eyes on me makes my every endeavor fearless and more meaningful: this thesis is dedicated to them.

Table of Contents

List of Tables

List of Figures

Abstract

Multiple access stability and broadcast delay in wireless networks

Nan Xie

Steven Weber, Ph.D.

This thesis addresses issues of design and performance analysis in wireless communication networks. We investigate topics relevant to both uplink and downlink. For uplink, we study the stability region of the slotted Aloha protocol under the collision channel model, for the case of a finite number of independent users. The stability region (i.e., the set of arrival rate vectors such that the whole queueing system can be made stable) is in general unknown when the number of users is more than two. We seek to characterize the set of *stabilizable* rate vectors, whereas most existing works only provide bounds on the region of stabilized rate vectors under a *given* control (i.e., a vector of contention probabilities). We choose a natural and important inner bound on the exact Aloha stability region. The results we obtain include equivalent forms of and alternative membership testing for this set, as well as other properties such as various geometrically intuitive and explicit inner and outer bounds, and generalized convexity properties of the associated "excess rate" functions.

For downlink, we seek to characterize the delay when broadcasting (random linear combinations of) information packets over independent erasure channels to a finite number of users. Of interest is the random delay until all the receivers recover the fixed chunk of packets initially queued at the base station (i.e., the sender). This falls into the study of certain order statistic of random variables. We obtain lower and upper bounds, exact expressions and a finite-step computational procedure (recurrence) for the moments of the random delay. We also investigate the dependence of the delay on the code blocklength (under random linear combinations of packets as the scheme employed in random linear network coding), and on the number of receivers, respectively. Results here include asymptotics, monotonicity properties and asymptotically tight lower and upper bounds.

Chapter 1: Introduction

1.1 Motivation and overview

As wireless communication networks keep growing and evolving, especially with the advent of high data rate (HDR) communications such as real-time multimedia networking, there is also an increasing demand for improving their performance, most notably, the *delay*. Smaller delay is always desirable, whether it is for uplink or downlink connections. This thesis is an effort towards meeting this challenge in the design and performance optimization of a wireless network. We seek to obtain an improved understanding of some performance metrics and their dependence on design parameters and strategies. Specifically we abstract a (one-hop) wireless network as a queueing system, focusing on the delay aspects. In particular, the stability problem studied in Chapter 2 is essentially about whether or not the delay is finite, and in Chapter 3 several properties of the (downlink) random broadcast delay including its moment expressions, asymptotics and non-asymptotic bounds, and its dependence on other key parameters in the model are investigated.

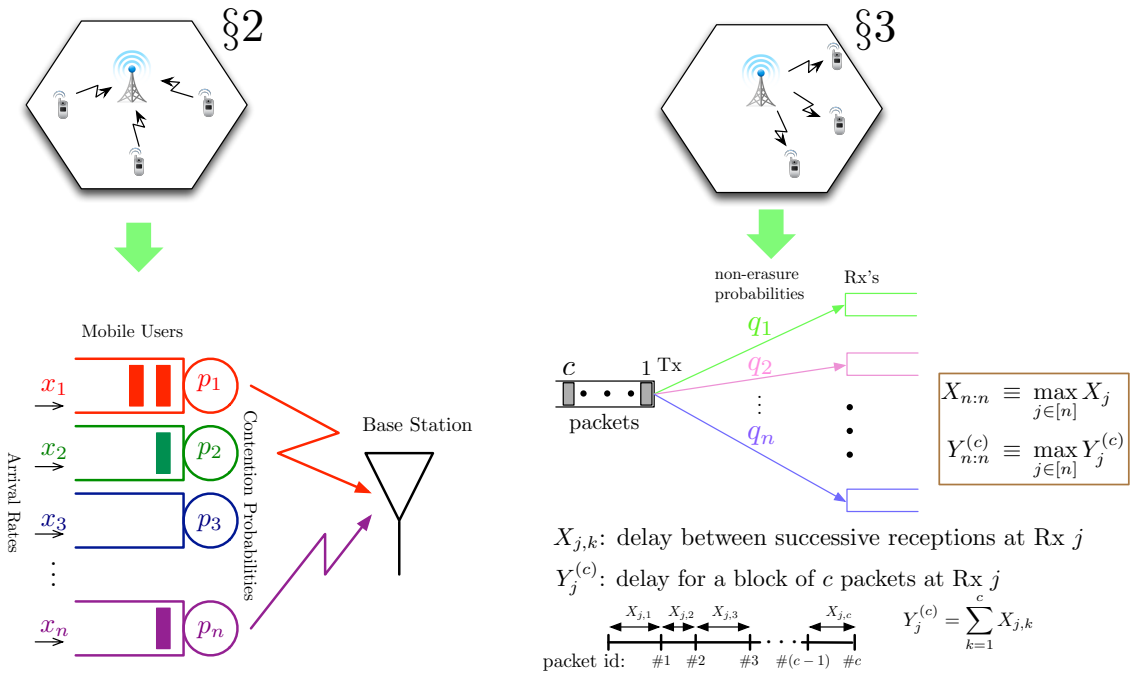

Figure 1.1: Overview of the thesis. **Left:** Chapter 2 on stability of the Aloha medium access control protocol. **Right:** Chapter 3 on broadcast delay over erasure channels.

Fig. 1.1 gives an overview of this thesis. The models will be detailed in subsequent chapters. For now it suffices to motivate the problems at an intuitive level. In particular, Chapter 2 is about the stability region of finite-user slotted-time Aloha medium access control protocol under the collision channel model. This model is illustrated in the left part of Fig. 1.1. The stability region is the set of arrival rate vectors $\mathbf{x}$ such that there exists some control (here, a vector of contention probabilities $\mathbf{p}$) under which no user will have its queue length (hence queueing delay) grow without bound. Two points need to be made clear. First, the control $\mathbf{p}$ is important from an operational viewpoint: if users all tend to choose large contention probabilities, implying at every time slot there is a very high likelihood for a collision to occur, then this is no good for any user in the system, and consequently the system is likely to be unstable as frequent collisions prevent each user's queue from being emptied infinitely often; if, on the contrary, every user tends to be too conservative, implying any given time slot is likely to be idle, then this is no good either, because the low service rates cannot sustain users' arrived packet streams, and consequently (some) users' queues will likely to grow to infinity, too. Second, regardless of its operational importance, from the perspective of testing membership in the stability region, this control parameter $\mathbf{p}$ is secondary because the stability problem is inherently about the existence of, rather than the exact value of, an appropriate control. If an arrival rate vector $\mathbf{x}$ is in the stability region then there must exist some control $\mathbf{p}$ that can stabilize this $\mathbf{x}$; if $\mathbf{x}$ is not in the stability region then there does not exist any control $\mathbf{p}$ that can stabilize $\mathbf{x}$. Note in this model, the support of $\mathbf{p}$ is $[0,1]^n$, an uncountably infinite set.

The paradigmatic set to be studied in Chapter 2 is

$$\Lambda \equiv \left\{ \mathbf{x} \in \mathbb{R}_+^n : \exists \mathbf{p} \in [0,1]^n : \ x_i \leq p_i \prod_{j \neq i} (1 - p_j), \ \forall i \in \{1, \ldots, n\} \right\},$$

which is a natural inner bound on the exact Aloha stability region. To see this, recall a fundamental result in queueing theory states that a queue is stable if the arrival rate λ is less than the service rate μ. In our model, the arrival rate for user i is given as an input parameter (i.e., x_i), the service rate is interpreted as the probability that this user attempts to contend for the channel while every other user does not, either because it elects not to or because its queue is empty (hence ineligible to

contend). Thus Λ is the set of all arrival rate vectors for which there exists some choice of contention probabilities such that the associated worst-case service rate (i.e., $\mu_i = p_i \prod_{j \neq i}(1 - p_j)$) for each user exceeds the user's arrival rate.

Next, Chapter 3 is about the broadcast delay of transmitting a block of c packets to n receivers over independent erasure channels, as shown in the right part of Fig. 1.1. Again, time is slotted. Because of the erasure channel and independence assumptions, the delay between successive receptions at a given receiver j, denoted $X_{j,k}$ (for $k \in \{1, \ldots, c\}$), is a geometric random variable (RV), and the delay for broadcasting a block of c packets to receiver j, denoted $Y_j^{(c)} \equiv \sum_{k=1}^{c} X_{j,k}$, is a (generalized) negative binomial RV. It then follows that the central objects of study throughout this chapter are the maximum order statistic of independent geometric RVs (denoted $X_{n:n}$), and of independent (generalized) negative binomial RVs (denoted $Y_{n:n}^{(c)}$).

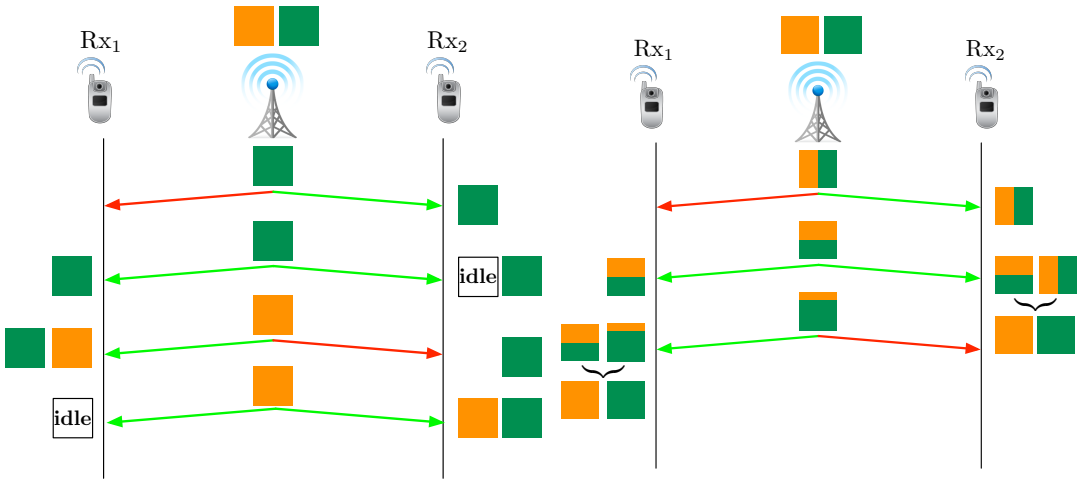

Figure 1.2: Illustration of why block-coding may be beneficial. **Left:** under UT (uncoded transmission), the total delay is $X_{2:2,1} + X_{2:2,2} = 4$. **Right:** under RLC (random linear combinations/coding), the total delay is $Y_{2:2}^{(2)} = 3$.

At this point a natural question is why we consider block-coding. Fig. 1.2 conveys intuition as to why block-coding may be beneficial. As shown, there are a total of $c = 2$ packets to be broadcast to two independent receivers (mobile users). For the same sequence of erasure channel realizations, random linear coding (RLC) completes the task with a total delay of $Y_{2:2}^{(2)} = 3$ slots, whereas under UT (uncoded transmission, i.e., serial transmission of uncoded packets) it takes $X_{2:2,1} + X_{2:2,2} = 4$ slots (or even more, if the realization for receiver 2's channel is an erasure during the 4^{th} slot

and onward). Here, random linear combinations/coding-based scheme means as long as a receiver obtains enough (here, c) linearly independent encoded packets it will be able to decode to retrieve the original c information packets. An important feature is that different receivers could achieve successful decoding from different combinations of encoded packets (as graphically illustrated in the figure, Rx's 1 and 2 receive different sets of encoded packets): this is a key to the reduction of some idle slots (as compared to UT). Essentially, under RLC, different receivers' reception processes are less coupled, resulting in fewer unnecessary wait slots from those receivers that are finished earlier.

1.2 Organization and contributions

We now highlight the main results obtained in the two major chapters: Chapter 2 and Chapter 3.

In Chapter 2 we address the uplink, focusing on the stability region of the finite-user slotted-time Aloha medium access control protocol. A well-known inner bound on the stability region is the set of all arrival rates for which there exists some choice of contention probabilities such that the associated worst-case service rate for each user exceeds the user's arrival rate, denoted Λ. Although testing membership in Λ of a given arrival rate vector can be posed as a convex program (Cor. 3), it is nonetheless of interest to understand the (geometric) properties of this set (and the set of stabilizing controls). Our first set of results (§2.2) hinges on a polynomial root testing, which not only provides an equivalent way of testing membership for Λ, but also helps establish some equivalent forms of Λ. These are indispensable later on when various optimization problems are formulated, reparameterized and solved. Furthermore, the augmented root testing allows one to find the critical stabilizing control(s) (shown to have intimate connection to the positive roots of a certain polynomial equation). Our second set of results (§2.4, §2.5, and §2.6) is about non-parametric bounds on Λ including simple and geometrically intuitive inner and outer bounds each induced by using a polyhedron, or a sphere, or an ellipsoid. Our last set of results (§2.7) is about the generalized convexity properties of natural "excess rate" functions associated with Λ, which have close connections to the set of stabilizing controls $\mathcal{P}(\mathbf{x})$. Many proofs of our results boil down to establishing certain inequalities for which the theory of convex sets and constrained optimization are essential. The interplay between algebra and geometry is demonstrated to be crucial too.

In Chapter 3 we turn to the downlink scenario where the random delay to broadcast a collection of c packets to n independent receivers is of interest. More specifically, we consider a transmitter broadcasting random linear combinations (over a finite or infinite field of size d) formed from a block of c packets to a total of n receivers, where the channels between the transmitter and each receiver are independent erasure channels with reception (i.e., non-erasure) probabilities $\mathbf{q} = (q_1, \ldots, q_n)$. We establish several properties of the random delay until all n receivers have recovered all c packets, denoted $Y_{n:n}^{(c)}$. First, we provide (§3.3 and §3.4) lower and upper bounds, exact expressions, and a recurrence for arbitrary (finite) moment of $Y_{n:n}^{(c)}$. Second, we study (§3.5) the dependence of per packet delay $Y_{n:n}^{(c)}/c$ on blocklength c including i) its convergence in probability and in r^{th} mean (as $c \to \infty$), ii) monotonicity of expected per packet delay, iii) asymptotically tight (in c) lower and upper moment bounds, and iv) a necessary and sufficient sample path monotonicity condition. Third, we employ extreme value theory to investigate (§3.6) the dependence of delay on the number of receivers n including the scaling (in n) of arbitrary (finite) moment of delay and asymptotically tight (in n) lower and upper moment bounds. Several results are new, some are extensions of existing results, and some are proofs of known results using new proof techniques. Throughout, a key tool we use is the notion of *stochastic ordering* which naturally bridges discrete RVs and their "continuous analogs" and helps establish bounds on arbitrary (finite) moment of the maximum order statistic of the discrete RVs of interest. We also make good use of extreme order statistic inequalities due to Ross and Peköz[2], and de la Cal and Cárcamo[3].

Chapter 2: Properties of an Aloha-like stability region

2.1 Introduction

2.1.1 Motivation and problem statement

This chapter addresses membership testing and structural properties of a natural inner bound on the stability region of the finite-user slotted-time Aloha medium access control (MAC) protocol under the collision channel model, hereafter *the Aloha protocol*[4]. The Aloha protocol is specified by a tuple $(n, \mathbf{x}, \mathbf{p})$ where $n \in \mathbb{N}$ is the number of users wishing to communicate with a common base station, $\mathbf{x} \in \mathbb{R}_+^n$ denotes the arrival rate of new packets at each user's queue (one queue per user, each queue assumed capable to hold an unlimited number of packets awaiting transmission), and $\mathbf{p} \in [0,1]^n$ denotes each user's chosen contention probability, i.e., the probability with which any user with a non-empty queue will contend for the channel. User contention decisions are synchronized at the beginning of each time slot, and, conditioned on the queue lengths, the user contention decisions are independent across users and across time slots. Each packet transmission requires exactly one time slot. Under the assumed collision channel model, an attempted transmission succeeds in a given time slot if and only if it is the only attempt in that time slot. Ternary channel feedback (success, collision, idle) from the base station to each user at the end of each time slot is assumed to be both instantaneous and error-free.

The stability region (of a MAC protocol) is defined as the set of arrival rate vectors (with elements corresponding to exogenous arrival rates at each user's queue) such that by some appropriate choice of the parameter(s) no user will, as time tends to infinity, accumulate an infinite backlog of packets waiting to be transmitted. The stability region asks for *necessary and sufficient* conditions in order for every user's queue to remain bounded. Let $q_i(t)$ denote user i's queue length at time t; queue i is stable if $\lim_{L\to\infty} \lim_{t\to\infty} \mathbb{P}(q_i(t) < L) = 1$, and the system is considered stable if every queue is stable. Since all the states of the underlying discrete time Markov chain (DTMC) of queue length vectors (defined on $\mathbb{Z}_+^n$) communicate, the stability of the system, or equivalently, the positive

recurrence of the DTMC amounts to the property that each queue has a non-zero probability of being empty, i.e., $\lim_{t\to\infty} \mathbb{P}(q_i(t) = 0) > 0$ for all $i \in \{1, \ldots, n\}$.

There is a significant body of work that derives bounds on the stability region of the Aloha protocol (denoted Λ_{A}) from a queueing-theoretic perspective, including[5–7]. In contrast to this approach, in this work we develop bounds and properties for an important and natural inner bound on the Aloha stability region, namely, the set of arrival rates for which there exists a vector of contention probabilities with associated worst-case service rates component-wise exceeding each arrival rate, denoted below by Λ. Our motivation to study this inner bound Λ is that testing membership of a candidate arrival rate vector $\mathbf{x}$ in Λ is easier than (but nonetheless has certain challenges similar to those encountered in) testing membership in the Aloha stability region. In either case one must, either implicitly or explicitly, identify $\mathbf{p}$, a vector of stabilizing contention probabilities, for which $\mathbf{x}$ can be shown to be in Λ or Λ_{A}. The difficulty is that the set of potential controls $\mathbf{p}$ is uncountably infinite ($\mathbf{p} \in [0, 1]^n$), and as such, given $\mathbf{x}$, it is not obvious whether or not such a $\mathbf{p}$ exists, i.e., whether or not $\mathbf{x}$ is stabilizable.

Our results address this challenge in several ways. First, we give a novel characterization of membership in Λ in terms of whether or not a certain order-n polynomial equation has a positive root. Second, we give several equivalent formulations for Λ, each with its own advantages. Third, we give a means of constructing a critical control $\mathbf{p}$ for any stabilizable rate vector $\mathbf{x}$. Fourth, we give polyhedral, spherical, and ellipsoid inner and outer non-parametric (explicit) bounds on Λ, which constitute, variously, necessary or sufficient conditions on membership in Λ. These explicit inner and outer bounds partially illuminate the shape and structure of Λ as a function of n. Finally, we present certain structural properties of certain functions and sets naturally associated with Λ, including the excess rate function and (an inner bound of) the set of contention probabilities that stabilize a given arrival rate vector.

The inner bound $\Lambda \subseteq \Lambda_{\mathrm{A}} \subseteq \mathbb{R}^n_+$ studied in this chapter is:

$$\Lambda \equiv \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \exists \mathbf{p} \in [0, 1]^n : \; x_i \leq p_i \prod_{j \neq i} (1 - p_j), \; \forall i \in \{1, \ldots, n\} \right\}. \tag{2.1}$$

The expression $p_i \prod_{j \neq i}(1-p_j)$ is the worst-case service rate for user i's queue, namely the service rate assuming all users have non-empty queues and thus all users are eligible for channel contention. In particular, user i's transmission is successful in such a time slot if user i elects to contend (with probability p_i) and each other user $j \neq i$ does not contend (each with independent probability $1-p_j$ for a non-empty queue). Clearly $\Lambda \subseteq \Lambda_{\mathrm{A}}$, since an arrival rate that is stabilizable under the worst-case service rate is certainly stabilizable under a better service rate. Our aim in this chapter is to establish properties of and non-parametric bounds on Λ. We emphasize that we call sets such as Λ "parametric" due to the observation that asserting membership in them requires explicitly or implicitly identifying another parameter which may be viewed as auxiliary from the perspective of membership testing.

In this chapter all the vectors are column vectors and inequalities between two vectors are understood to hold component-wise. A list of general notation is given in Table 3.3.

Table 2.1: General notation (for Chapter 2)

Symbol	Meaning
n	number of users, also the default length of a vector
$[n] = \{1, \ldots, n\}$	set of positive integers up to n
$\mathbf{x} = (x_1, \ldots, x_n)$	vector of user's arrival rates
$\mathbf{p} = (p_1, \ldots, p_n)$	vector of user's chosen probabilities for channel contention
$\mathbf{1}$	a vector with 1 in all its positions
$\mathbf{e}_i$	unit vector with 1 in position $i \in [n]$
$\mathbf{m} = m\mathbf{1}$, $m = \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}$	"all-rates-equal" point $\mathbf{m}$ and its component value m
$\pi(\mathbf{p}) = \prod_i (1-p_i)$	product of $(1-p_i)$'s
$\mathbf{x}(\mathbf{p})$ for $x_i(\mathbf{p}) = p_i \prod_{j \neq i}(1-p_j)$	n-vector $\mathbf{x}(\mathbf{p})$ with components $x_i(\mathbf{p})$ determined by $\mathbf{p}$
$\mathbf{p}(\delta, \mathbf{x})$ for $p_i(\delta, \mathbf{x}) = \frac{\delta x_i}{1+\delta x_i}$	n-vector $\mathbf{p}(\delta, \mathbf{x})$ with $p_i(\delta, \mathbf{x})$ determined by $\mathbf{x}$ and parameterized by $\delta > 0$
bd	topological boundary of a set
int	interior of a set
conv	convex hull of a set
$\overline{A}$	closure of set A
A^c	complement of set A
$\|\cdot\|$	l_2 norm
$d(x,y) = \|x-y\|$	Euclidean distance between (geometric objects) x, y
$\mathbb{I}_{\mathrm{S}}$	indicator function for boolean expression S
$\mathcal{S} = \{\mathbf{z} \geq \mathbf{0} : \sum_i z_i \leq 1\}$	closed standard unit simplex
$\partial\mathcal{S} = \{\mathbf{z} \geq \mathbf{0} : \sum_i z_i = 1\}$	the facet of $\mathcal{S}$, namely the set of probability vectors
$\mathcal{H}(\mathbf{n}, d) = \{\mathbf{x} : \mathbf{n}^{\mathsf{T}}\mathbf{x} = d\}$	hyperplane with normal vector $\mathbf{n}$ and displacement d
$\mathcal{B}(\mathbf{c}, r) = \{\mathbf{x} : d(\mathbf{c}, \mathbf{x}) < r\}$	open ball centered at $\mathbf{c}$ with radius r
$\mathcal{E} = \{\mathbf{x} : (\mathbf{x}-\mathbf{c})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{x}-\mathbf{c}) < 1\}$	open ellipsoid centered at $\mathbf{c}$, expressed in quadratic form
$\mathcal{E}(c, a_1, a_2)$	open ellipsoid centered at $c\mathbf{1}$ with semi-axis lengths a_1, $a_2 = \cdots = a_n$
$\mathbf{Q}$	the rotation matrix used in Prop. 12, also related to ellipsoid's axes (§2.6)

2.1.2 Related work

The throughput analysis of the Aloha packet system with and without slots can be found in Roberts[8] and Abramson[9]. The Aloha stability region problem was posed in 1979 by Tsybakov and Mikhailov[10] who also solved the $n = 2$ and the homogeneous n-user case, for both of which they showed $\Lambda = \Lambda_{\mathrm{A}}$. Szpankowski[11] studied this problem when $n > 2$, with result expressed in terms of the joint statistics of the queue lengths. The use of the so-called "dominant system" in Rao and Ephremides[5], as well as Luo and Ephremides[6], established some important bounds on the stability region. Anantharam[12] showed $\Lambda = \Lambda_{\mathrm{A}}$ for a certain correlated arrival process by applying the Harris correlation inequality. Using mean field analysis, assuming each queue's evolution is independent, Bordenave et al.[13] were able to show $\Lambda = \Lambda_{\mathrm{A}}$ holds asymptotically in n. Recently Kompalli and Mazumdar[7] obtained bounds that are linear with respect to the users' arrival rates, based on a Foster-Lyapunov approach. To date, characterization of Λ_{A} remains open for the general n-user case with general arrival processes, although it's been conjectured ([5] §V,[14] §V Thm. 2) that Λ coincides with the Aloha stability region Λ_{A}. More recently, Subramanian and Leith[15] showed structural properties such as boundary and convexity properties of the rate region of CSMA/CA wireless local-area networks which includes Aloha and IEEE 802.11 as special cases.

Besides its intimate connection with the Aloha stability region Λ_{A}, the set Λ has also been featured in an information theoretic context. Namely, in 1985 Massey and Mathys[16] proved Λ is the capacity region of the collision channel without feedback. In the same issue, Post[17] established the convexity of the complement of Λ in the non-negative orthant $\mathbb{R}^n_+$.

2.1.3 Summary of bounds on Λ

In this chapter we present a variety of inner and outer bounds on Λ, including the "square-root-sum" inner bound Λ_{srs} (§2.3, Prop. 5), polyhedral inner Λ_{pi} and outer Λ_{po} bounds (§2.4, Props. 7 and 8), spherical inner Λ_{si} and outer Λ_{so} bounds (§2.5, Props. 9 and 10), and ellipsoid inner Λ_{ei} and outer Λ_{eo} bounds (§2.6, Props. 18 and 17). The volumes of the aforementioned bounds as a function of the number of users, n, are collected in Fig. 2.1 and Tables 2.2 and 2.3, where i (o) refers to inner (outer) bound respectively, and p, s, e refers to polyhedral, spherical, ellipsoid, respectively. We give

the volumes themselves, as well as the volumes normalized by the volume of the (trivial) simplex outer bound of $1/n!$. The volumes of $\Lambda_{\rm srs}$, $\Lambda_{\rm pi}$, Λ are computed exactly from closed-form expressions we derive in the chapter. All other volumes are estimated using standard Monte-Carlo simulation. $\Lambda^*_{\rm pi}$ is the optimal polyhedral inner bound among its family. For the spherical bounds $\Lambda_{\rm si}$, $\Lambda_{\rm so}$ the center of spheres are chosen such that the induced bounds are optimal within their families (hence the notation $\Lambda^*_{\rm si}$ and $\Lambda^*_{\rm so}$). For the ellipsoid bounds $\Lambda_{\rm ei}$, $\Lambda_{\rm eo}$ the center of ellipsoids are chosen by setting $c = 2$. $\Lambda_{\rm po\&so}$ is constructed by using $\Lambda_{\rm po}$ and $\Lambda^*_{\rm so}$ in conjunction namely $\Lambda_{\rm po\&so} \equiv \Lambda_{\rm po} \cap \Lambda^*_{\rm so}$. It is clear from Fig. 2.1 that the three inner bounds (polyhedral, spherical, ellipsoid) are tighter than are the four outer bounds.

For the Monte-Carlo volume estimates we generate independent points over $[0,1]^n$ uniformly at random, and use the fraction of points that fall into the region defined by the bound as our volume estimate. As the volume of the unit box $[0,1]^n$ is 1, the volume of any subset of the unit box can be viewed as the probability that a point uniformly distributed over the unit box falls into this subset, which equals the mean of a Bernoulli random variable, say $Z \sim \text{Ber}(v)$, for v the volume of the subset. This justifies the use of the sample mean, $\hat{v}_k = (Z_1 + \cdots + Z_k)/k$, as our volume estimate. We also include confidence interval estimates in both tables, indicating the relative half-width, denoted δ, in order for the probability that the true mean v deviates from the sample mean $\hat{v}_k$ by a fraction of no more than δ is at least $1-\alpha$. More precisely, let n and the bound (with unknown volume v) be given, and let k be the total number of trials for generating instances of i.i.d. random variables $Z \sim \text{Ber}(v)$. We want to find δ such that $\mathbb{P}\left(\hat{v}_k\left(1-\delta\right) \le v \le \hat{v}_k\left(1+\delta\right)\right) \ge 1-\alpha$. Under a normal approximation we can derive $\delta \approx \sqrt{\frac{1-\hat{v}_k}{(k-1)\hat{v}_k}}\Phi^{-1}(1-\alpha/2)$, applying results from[18] (§9.1). In our simulations we use $k = 10^8$ and $\alpha = 5\%$. For those volumes estimated using Monte-Carlo, the corresponding entries in Table 2.2 are $\hat{v}_k$ (top) and δ (bottom), and in Table 2.3 are $n!\hat{v}_k$ (top) and δ (bottom).

As will be shown (Remarks 1, 3 and 5), the inner bounds are ordered by volume for all $n \ge 3$[1],

[1]Provided the inner bounding ellipsoid is such that its center $\mathbf{c} = c\mathbf{1}$ with $c \ge (1-nm^2)/(2(1-nm))$ where $m = m(n) \equiv \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}$.

i.e.,

$$\mathrm{vol}(\Lambda_{\mathrm{srs}}) \leq \mathrm{vol}(\Lambda^*_{\mathrm{pi}}) \leq \mathrm{vol}(\Lambda^*_{\mathrm{si}}) \leq \mathrm{vol}(\Lambda_{\mathrm{ei}}) \leq \mathrm{vol}(\Lambda),\ n \geq 3. \tag{2.2}$$

Among the outer bounds (Λ_{eo}, $\Lambda_{\mathrm{po\&so}}$, Λ_{po} and Λ^*_{so}) there is no such complete ordering valid for all n, although $\Lambda_{\mathrm{po\&so}}$ outperforms both Λ_{po} and Λ^*_{so} by construction, and the ellipsoid outer bound Λ_{eo} outperforms the optimal spherical outer bound Λ^*_{so} provided the outer bounding ellipsoid is such that its center $\mathbf{c} = c\mathbf{1}$ with $c \geq 1$.

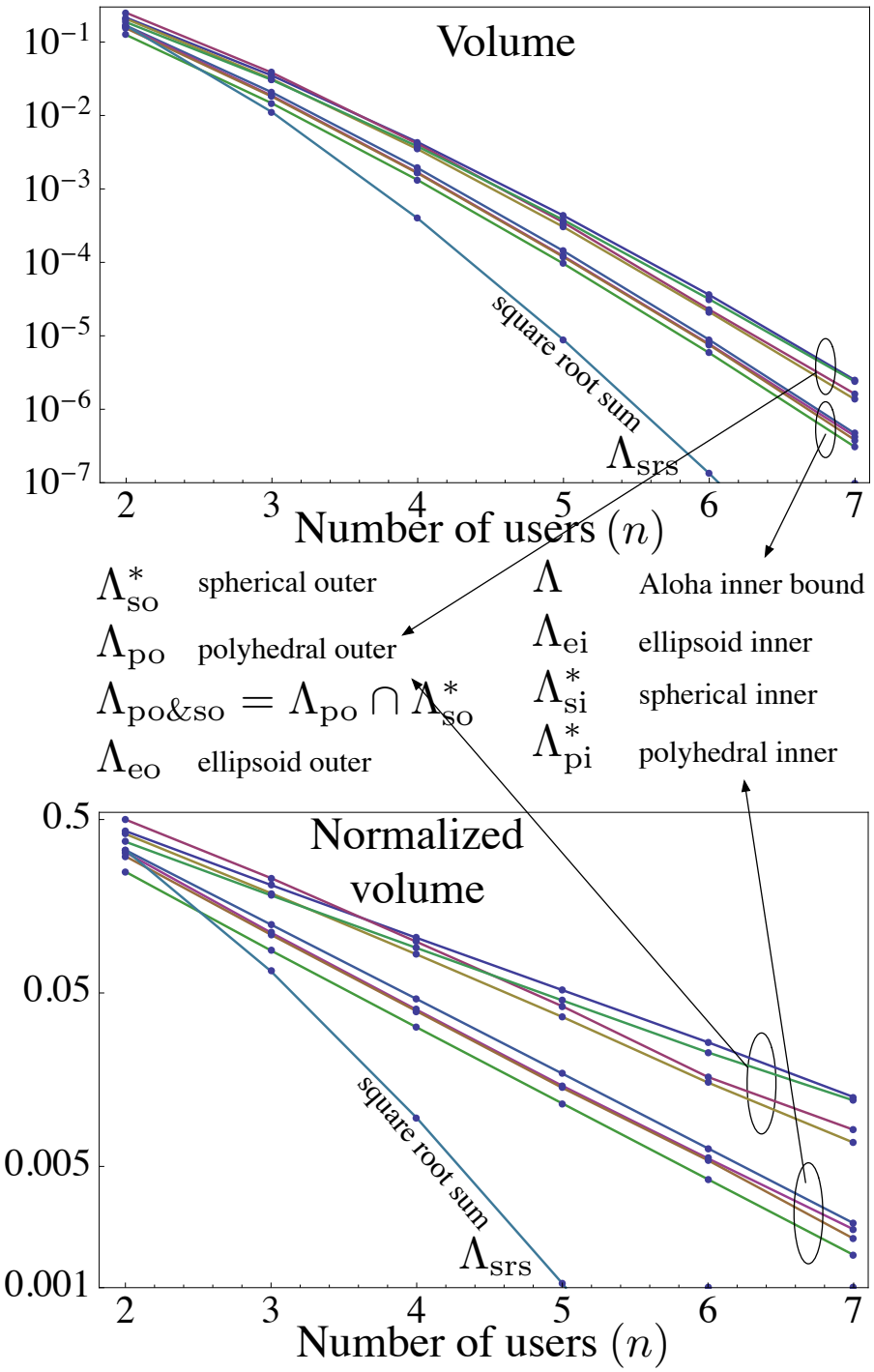

Figure 2.1: The volumes (computed, or estimated from Monte-Carlo simulation) of the various inner and outer bounds $\Lambda_{\text{srs}}, \Lambda^*_{\text{pi}}, \Lambda^*_{\text{si}}, \Lambda_{\text{ei}}, \Lambda_{\text{eo}}, \Lambda_{\text{po\&so}}, \Lambda_{\text{po}}, \Lambda^*_{\text{so}}$ on the Aloha stability region inner bound Λ versus the number of users, n. The top figure shows the volumes and the bottom figure shows the volumes normalized by the volume of the (trivial) simplex outer bound $(1/n!)$. Each of the two ovals on each plot groups four curves, with the top oval indicating the four left labels and the bottom oval indicating the four right labels. The left and right label orderings reflect the ordering of the curves within the top and bottom ovals, respectively. Ellipsoids have parameter $c = 2$.

Table 2.2: Volumes (computed or estimated) of the various bounds (estimates include normalized 95% CI, δ)

Bounds (§)	$n=2$ ($\times 10^{-1}$)	$n=3$ ($\times 10^{-2}$)	$n=4$ ($\times 10^{-3}$)	$n=5$ ($\times 10^{-4}$)	$n=6$ ($\times 10^{-6}$)	$n=7$ ($\times 10^{-7}$)
Spherical outer bound Λ^*_{so} (§2.5)	2.146	3.491	4.330	4.328	36.09	24.900
	0.0004	0.0010	0.0030	0.0094	0.0326	0.1242
Polyhedral outer bound Λ_{po} (§2.4)	2.500	3.828	4.106	3.492	22.75	16.127
	0.0003	0.0010	0.0031	0.0105	0.0411	0.1543
Polyhedral and spherical outer bound $\Lambda_{\text{po\&so}}$	2.064	3.132	3.475	3.031	21.14	13.600
	0.0004	0.0011	0.0033	0.0113	0.0426	0.1681
Ellipsoid outer bound Λ_{eo} (§2.6)	1.862	3.044	3.785	3.774	31.53	23.800
	0.0004	0.0011	0.0032	0.0100	0.0349	0.1270
Aloha stability region inner bound $\boldsymbol{\Lambda}$ (§2.3)	**1.667**	**2.063**	**1.921**	**1.428**	**8.821**	**4.665**
Ellipsoid inner bound Λ_{ei} (§2.6)	1.618	1.856	1.667	1.214	7.720	4.300
	0.0004	0.0014	0.0048	0.0178	0.0705	0.2989
Spherical inner bound Λ^*_{si} (§2.5)	1.535	1.805	1.635	1.183	7.530	3.800
	0.0005	0.0014	0.0048	0.0180	0.0714	0.3179
Polyhedral inner bound Λ^*_{pi} (§2.4)	1.250	1.463	1.320	0.961	5.851	3.061
Square root sum inner bound Λ_{srs} (§2.3)	1.667	1.111	0.397	0.0882	0.134	0.0147

Table 2.3: Normalized volumes (by $1/n!$) of the various bounds (estimates include normalized 95% CI, δ)

Bounds (§)	$n=2$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
Spherical outer bound Λ^*_{so} (§2.5)	0.4291	0.2095	0.1039	0.0519	0.0260	0.01255
	0.0004	0.0010	0.0030	0.0094	0.0326	0.1242
Polyhedral outer bound Λ_{po} (§2.4)	0.5000	0.2297	0.0985	0.0419	0.0164	0.0081
	0.0003	0.0010	0.0031	0.0105	0.0411	0.1543
Polyhedral and spherical outer bound $\Lambda_{\text{po\&so}}$	0.4129	0.1879	0.0834	0.0364	0.0152	0.0069
	0.0004	0.0011	0.0033	0.0113	0.0426	0.1681
Ellipsoid outer bound Λ_{eo} (§2.6)	0.3724	0.1827	0.0908	0.0453	0.0227	0.0120
	0.0004	0.0011	0.0032	0.0100	0.0349	0.1270
Aloha stability region inner bound $\boldsymbol{\Lambda}$ (§2.3)	**0.3333**	**0.1234**	**0.0461**	**0.0171**	**0.0064**	**0.0024**
Ellipsoid inner bound Λ_{ei} (§2.6)	0.3236	0.1113	0.0400	0.0146	0.0056	0.0022
	0.0004	0.0014	0.0048	0.0178	0.0705	0.2989
Spherical inner bound Λ^*_{si} (§2.5)	0.3070	0.1083	0.0392	0.0142	0.0054	0.0019
	0.0005	0.0014	0.0048	0.0180	0.0714	0.3179
Polyhedral inner bound Λ^*_{pi} (§2.4)	0.2500	0.0878	0.0317	0.0116	0.0042	0.0015
Square root sum inner bound Λ_{srs} (§2.3)	0.3333	0.0667	0.0095	0.0011	0.0001	0.000007

2.1.4 Organization and contributions

We now describe the major sections of the chapter, highlighting our main results in each section. In §2.2 we present a polynomial root condition for testing membership in Λ, and use this result to establish some equivalent forms of Λ. Furthermore, the root testing can be augmented so that it allows us to exclusively find the critical stabilizing control(s). In §2.3 we compute the volume of Λ

in closed-form, meaning it is expressed as a finite (albeit complicated) sum. We then give a simple inner bound on Λ, exact for $n = 2$, but provably quite weak for $n > 2$. The next three sections give explicit (non-parametric) inner and outer bounds on Λ. Specifically, §2.4 gives the optimal polyhedral inner bound induced by a single hyperplane as well as a polyhedral outer bound in $\mathbb{R}^n_+$ induced by $n + 1$ hyperplanes, §2.5 presents the optimal spherical inner and outer bounds each induced by a single sphere, and §2.6 establishes ellipsoid inner and outer bounds each induced by an ellipsoid. We include two different proofs for our ellipsoid outer bound. Our last technical section, §2.7, shifts the focus to the generalized convexity properties of an "excess rate" function associated with Λ, and establishes the convexity of the set of stabilizing controls for a given rate vector assuming worst-case service rate. Several of the proofs are placed in an appendix that follows.

2.2 Polynomial membership testing, forms of Λ, and critical stabilizing controls

This section introduces three rather distinct results which are presented together on account of the fact that their proofs rely upon closely related concepts. First, Prop. 1 demonstrates that testing membership of a rate vector $\mathbf{x}$ in Λ is equivalent to a certain polynomial equation having at least one positive root. Second, Prop. 2 establishes two set definitions similar to Λ are in fact equivalent to Λ. Finally, Prop. 3 identifies a "critical stabilizing control" $\mathbf{p}(\mathbf{x})$ (see Def. 3) for each $\mathbf{x} \in \Lambda$. Def. 1 gives three sets, related to Λ, that will be important for what follows.

Definition 1.

$$\Lambda_{\text{eq}} \equiv \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \exists \mathbf{p} \in [0,1]^n : \ x_i = p_i \prod_{j \neq i} (1 - p_j), \ \forall i \in \{1, \ldots, n\} \right\} \tag{2.3}$$

$$\Lambda_{\partial \mathcal{S}} \equiv \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \exists \mathbf{p} \in [0,1]^n, \sum_i p_i = 1 : \ x_i \leq p_i \prod_{j \neq i} (1 - p_j), \ \forall i \in \{1, \ldots, n\} \right\} \tag{2.4}$$

$$\partial \Lambda \equiv \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \exists \mathbf{p} \in [0,1]^n, \sum_i p_i = 1 : \ x_i = p_i \prod_{j \neq i} (1 - p_j), \ \forall i \in \{1, \ldots, n\} \right\}. \tag{2.5}$$

Comparison with Λ in (2.1) makes clear that Λ_{eq} replaces all the inequalities in Λ with equalities, $\Lambda_{\partial \mathcal{S}}$ adds to Λ a restriction that the contention probabilities sum to one, and $\partial \Lambda$ adds both of these

to Λ. We denote the set of all sub-stochastic vectors as $\mathcal{S} \equiv \{\mathbf{z} \geq \mathbf{0} : \sum_i z_i \leq 1\}$, and its facet in $\mathbb{R}^n_+$, the set of all stochastic vectors (also called probability vectors) as $\partial\mathcal{S} \equiv \{\mathbf{z} \geq \mathbf{0} : \sum_i z_i = 1\}$; this notation explains the label $\Lambda_{\partial\mathcal{S}}$. The next definition introduces several quantities to be used. Let $\{\mathbf{e}_i\}_{i=1}^n$ denote the n standard unit vectors in $\mathbb{R}^n_+$.

Definition 2. *The order-n polynomial in $\delta \in \mathbb{R}$ with coefficients determined by $\mathbf{x} \in [0,1]^n \setminus \{\mathbf{e}_i\}_{i=1}^n$:*

$$f(\delta, \mathbf{x}) \equiv \prod_{i=1}^{n}(1 + x_i\delta) - \delta. \tag{2.6}$$

The n-vector $\mathbf{p}(\delta, \mathbf{x}) \in [0,1]^n$ with components $p_i(\delta, \mathbf{x})$ determined by $\mathbf{x} \in \mathbb{R}^n_+$ and parameterized by $\delta > 0$:

$$p_i(\delta, \mathbf{x}) \equiv \frac{\delta x_i}{1 + \delta x_i}, i \in [n]. \tag{2.7}$$

The product of the component-wise complements of a given vector of contention probabilities $\mathbf{p} \in [0,1]^n$:

$$\pi(\mathbf{p}) \equiv \prod_j (1 - p_j). \tag{2.8}$$

The n-vector $\mathbf{x}(\mathbf{p}) \in [0,1]^n$ with components $x_i(\mathbf{p})$ determined by $\mathbf{p} \in [0,1]^n$:

$$x_i(\mathbf{p}) \equiv p_i \prod_{j \neq i}(1 - p_j) = \frac{p_i}{1 - p_i}\pi(\mathbf{p}),\ i \in [n]. \tag{2.9}$$

Note the equalities in Λ_{eq} in (2.3) are $\mathbf{x} = \mathbf{x}(\mathbf{p})$ in (2.9). Our notation distinguishes between a generic vector of contention probabilities $\mathbf{p} \in [0,1]^n$ and a specific vector $\mathbf{p}(\delta, \mathbf{x})$ determined by δ and $\mathbf{x}$, and likewise between a generic rate vector $\mathbf{x} \in \Lambda$ and a specific vector $\mathbf{x}(\mathbf{p})$ determined by $\mathbf{p}$.

Definition 3 (stabilizability in the sense of Λ or its equivalent forms)**.** *A stabilizing control for $\mathbf{x} \in \Lambda$ is a vector of contention probabilities $\mathbf{p} \in [0,1]^n$ that is "compatible" with $\mathbf{x}$, meaning the pair $(\mathbf{x}, \mathbf{p})$ satisfies the definition of Λ. A* critical stabilizing control *is a stabilizing control $\mathbf{p}$ such that $(\mathbf{x}, \mathbf{p})$ satisfies the definition of Λ_{eq} (or $\partial\Lambda$).*

A corollary of Prop. 1 below is that, given $\mathbf{x} \in [0,1]^n$, there exists a stabilizing control iff there exists a critical stabilizing control.

Since $\Lambda \subseteq \Lambda_{\rm A}$, the non-existence of a stabilizing control for a given $\mathbf{x}$ in the sense of Λ does not necessarily mean the non-existence of one for $\mathbf{x}$ in the sense of $\Lambda_{\rm A}$ (i.e., it does not necessarily mean $\mathbf{x}$ is not stabilizable under the Aloha protocol). Throughout this chapter though, our usage of "stabilizability" and "stability controls" is tied to Λ or its equivalent forms.

The following proposition gives an alternative test for membership of a rate vector $\mathbf{x}$ in Λ in terms of the existence of a positive root of the polynomial $f(\delta, \mathbf{x})$ in (2.6), and furthermore establishes that in fact $\Lambda = \Lambda_{\rm eq}$. The converse proof is constructive, meaning given a positive root δ, one can construct a $\mathbf{p}(\delta, \mathbf{x})$ compatible with $\mathbf{x}$. In the forward direction, given $\mathbf{x} \in \Lambda$ and an associated compatible $\mathbf{p}$, we do not give an explicit expression for a positive root δ of $f(\delta, \mathbf{x})$, although we can lower bound the interval containing δ. In the forward direction for $\mathbf{x} \in \Lambda_{\rm eq}$, however, given a compatible $\mathbf{p}$ such that $\mathbf{x} = \mathbf{x}(\mathbf{p})$ as in (2.9), we have that one of the positive roots of $f(\delta, \mathbf{x})$ for $\mathbf{x} \in \Lambda_{\rm eq}$ will always equal $\delta = 1/\pi(\mathbf{p})$.

Proposition 1 (root testing)**.** *Membership in Λ (except $\{\mathbf{e}_i\}_{i=1}^n$) is equivalent to the existence of a positive root of the polynomial equation $f(\delta, \mathbf{x}) = 0$.*

$$\mathbf{x} \in \Lambda \setminus \{\mathbf{e}_i\}_{i=1}^n \iff \exists \delta > 0 \ : \ f(\delta, \mathbf{x}) = 0. \tag{2.10}$$

The same equivalence holds for membership in $\Lambda_{\rm eq}$. In particular, $\Lambda = \Lambda_{\rm eq}$.

Proof. We first address membership testing for Λ.

"$\Leftarrow$": Fix $\mathbf{x} \notin \{\mathbf{e}_i\}_{i=1}^n$ and suppose $\delta > 0$ satisfies $f(\delta, \mathbf{x}) = 0$. Construct $\mathbf{p}(\delta, \mathbf{x})$ as in (2.7), and observe the worst-case service rate for user i is

$$p_i(\delta, \mathbf{x}) \prod_{j \neq i} (1 - p_j(\delta, \mathbf{x})) = \frac{\delta x_i}{\prod_j (1 + \delta x_j)}, \tag{2.11}$$

and for this choice of $\mathbf{p}$ the requirement $x_i \leq p_i \prod_{j \neq i}(1 - p_j)$ simplifies to $\prod_j (1 + \delta x_j) \leq \delta$, which is true with equality by assumption that $f(\delta, \mathbf{x}) = 0$. As this is true for each $i \in [n]$ it follows that $\mathbf{x} \in \Lambda$.

"$\Rightarrow$": First observe that if $p_i = 1$ for some $i \in [n]$ then the only way for $\mathbf{x} \in \Lambda$ is to let $\mathbf{x} \leq \mathbf{e}_i$,

implying $\mathbf{x} \in \Lambda$. Similarly, if $x_i = 0$ for some $i \in [n]$, then we can work with a reduced-dimensional $\mathbf{x}$ (i.e., the original $\mathbf{x}$ with zero component(s) removed). Consequently, we now assume $p_i < 1$ and $x_i > 0$ for each $i \in [n]$. Suppose $\mathbf{x} \in \Lambda \setminus \{\mathbf{e}_i\}_{i=1}^n$ and let $\mathbf{p}$ be compatible with $\mathbf{x}$. Define the "inverse stability rank" vector $\boldsymbol{\Delta}$ (Luo and Ephremides[6] Thm. 2) with elements

$$\Delta_i = \frac{p_i}{x_i(1-p_i)},\ i \in [n]. \tag{2.12}$$

Then $\mathbf{x} \in \Lambda$ may be equivalently expressed in terms of $\boldsymbol{\Delta}$ via:

$$\begin{aligned}
\mathbf{x} \in \Lambda \quad &\iff \quad \exists \mathbf{p} : x_i \le p_i \prod_{j \neq i}(1-p_j),\ i \in [n] \\
&\iff \quad \exists \mathbf{p} : \frac{p_i}{x_i(1-p_i)} \ge \prod_{j \in [n]} \frac{1}{1-p_j} = \prod_{j \in [n]} \left(1 + x_j \frac{p_j}{x_j\,(1-p_j)}\right),\ i \in [n] \\
&\iff \quad \exists \boldsymbol{\Delta} : \Delta_i \ge \prod_j (1 + x_j \Delta_j),\ i \in [n] \qquad (*)
\end{aligned} \tag{2.13}$$

Define $\tilde{\Delta} \equiv \min_j \Delta_j$, and let $\tilde{\boldsymbol{\Delta}} = \tilde{\Delta}\mathbf{1}$ be the n-vector with all components equal to $\tilde{\Delta}$. If $\boldsymbol{\Delta}$ obeys $(*)$ in (2.13) then $\tilde{\boldsymbol{\Delta}}$ also obeys $(*)$, because

$$\tilde{\Delta} = \tilde{\Delta}_i = \min_j \Delta_j \ge \prod_k (1 + x_k \Delta_k) \ge \prod_k (1 + x_k \tilde{\Delta}_k) = \prod_k (1 + x_k \tilde{\Delta}),\ i \in [n].$$

It follows that $f(\tilde{\Delta}, \mathbf{x}) \le 0$. If $f(\tilde{\Delta}, \mathbf{x}) = 0$ then $\tilde{\Delta}$ is the required positive root in (2.10). Otherwise, notice $\lim_{\tilde{\Delta} \to \infty} f(\tilde{\Delta}, \mathbf{x}) = \infty$, so by the intermediate value theorem there must exist some $\delta \in (\tilde{\Delta}, \infty)$ so that $f(\delta, \mathbf{x}) = 0$. This concludes the proof of the equivalence for membership testing for Λ.

We now address membership testing for Λ_{eq}.

"$\Leftarrow$": The same proof used for membership testing for Λ holds here.

"$\Rightarrow$": We must show that if $\mathbf{x} \in \Lambda_{\text{eq}} \setminus \{\mathbf{e}_i\}_{i=1}^n$ then there exists $\delta > 0$ such that $f(\delta, \mathbf{x}) = 0$. But the above proof for membership testing for Λ showed such a δ always exists for each $\mathbf{x} \in \Lambda$, and as $\Lambda_{\text{eq}} \subseteq \Lambda$, a δ must likewise exist for each $\mathbf{x} \in \Lambda_{\text{eq}}$. The fact that $f(1/\pi(\mathbf{p}), \mathbf{x}) = 0$ for $\mathbf{p}$ compatible with $\mathbf{x} \in \Lambda_{\text{eq}}$ follows by substitution. This concludes the proof of the equivalence for membership testing for Λ_{eq}.

The assertion $\Lambda = \Lambda_{\text{eq}}$ is immediate from the two equivalences just established. □

The following proposition extends the equivalence of Λ and Λ_{eq} to include $\Lambda_{\partial\mathcal{S}}$.

Proposition 2. $\Lambda_{\text{eq}} = \Lambda = \Lambda_{\partial\mathcal{S}}$.

Proof. Prop. 1 established $\Lambda_{\text{eq}} = \Lambda$; it remains to show $\Lambda = \Lambda_{\partial\mathcal{S}}$. As $\Lambda_{\partial\mathcal{S}} \subseteq \Lambda$, we only need to show $\Lambda \subseteq \Lambda_{\partial\mathcal{S}}$. By Lem. 1 of[16], given $\mathbf{x} \in \Lambda$ (with compatible $\mathbf{p} \in [0,1]^n$), there must exist a unique $\hat{\mathbf{x}} \in \partial\Lambda$ (with a unique compatible $\hat{\mathbf{p}} \in \partial\mathcal{S}$) which "dominates" $\mathbf{x}$ in the sense that $\mathbf{x} \leq \hat{\mathbf{x}}$. If in fact $\mathbf{x} = \hat{\mathbf{x}}$ then $\hat{\mathbf{p}} = \mathbf{p}$ as well[16]. Since $\hat{\mathbf{x}} \in \partial\Lambda$ and $\mathbf{x} \leq \hat{\mathbf{x}}$, it follows that $\mathbf{x} \in \Lambda_{\partial\mathcal{S}}$, and thus $\Lambda \subseteq \Lambda_{\partial\mathcal{S}}$. □

We next present an augmented version of the root testing Prop. 1, which makes clear how the roots of polynomial equation $f(\delta, \mathbf{x}) = 0$ map between compatible $\mathbf{p}$ and $\mathbf{x} \in \Lambda = \Lambda_{\text{eq}}$. The proofs of the (critical) stabilizing controls $\mathbf{p}(\mathbf{x})$ for a given $\mathbf{x}$ are constructive.

Proposition 3 (augmented root testing)**.** *Fix $n \geq 2$ and let a rate vector $\mathbf{x} \in [0,1]^n \setminus \{\mathbf{e}_i\}_{i=1}^n$ be given.*

1. *$\mathbf{x} \in \partial\Lambda$ iff there is a unique positive root δ of $f(\delta, \mathbf{x}) = 0$, denoted δ_u. Furthermore given $\mathbf{x} \in \partial\Lambda$, then $\mathbf{p}_u = \mathbf{p}(\delta_u, \mathbf{x})$ given by (2.7) stabilizes $\mathbf{x}$. Finally, $\mathbf{p}_u \in \partial\mathcal{S}$ and is the only (critical) stabilizing control for $\mathbf{x}$ among all $\mathbf{p} \in [0,1]^n$.*

2. *Let $\mathbf{x} \in \Lambda \setminus \partial\Lambda$ be given. Solving $f(\delta, \mathbf{x}) = 0$ on $(0, \infty)$ for δ yields exactly two positive roots denoted δ_s, δ_l. Each root can be used to construct a vector of contention probabilities, $\mathbf{p}_s = \mathbf{p}(\delta_s, \mathbf{x})$, $\mathbf{p}_l = \mathbf{p}(\delta_l, \mathbf{x})$, according to (2.7), that stabilizes $\mathbf{x}$. Furthermore, $\mathbf{p}_s$ is such that $\sum_{i=1}^n p_{s,i} < 1$ (i.e., $\mathbf{p}_s \in \mathcal{S} \setminus \partial\mathcal{S}$) and $\mathbf{p}_l$ is such that $\sum_{i=1}^n p_{l,i} > 1$ (i.e., $\mathbf{p}_l \in [0,1]^n \setminus \mathcal{S}$). Finally, $\mathbf{p}_s$, $\mathbf{p}_l$ are also the only two critical stabilizing controls for $\mathbf{x}$ among all $\mathbf{p} \in [0,1]^n$.*

Proof. See §2.8.1 in the Appendix. □

Corollary 1. *There exist the following bijections: i) $\partial\mathcal{S} \leftrightarrow \partial\Lambda$, ii) $\mathcal{S} \setminus \partial\mathcal{S} \leftrightarrow \Lambda \setminus \partial\Lambda$, iii) $[0,1]^n \setminus \mathcal{S} \leftrightarrow \Lambda \setminus \partial\Lambda$, iv) $\mathcal{S} \setminus \partial\mathcal{S} \leftrightarrow [0,1]^n \setminus \mathcal{S}$, v) $\mathcal{S} \leftrightarrow \Lambda$.*

Proof. Massey and Mathys[16] showed *i*). We now show *ii*). From (the proof of) Prop. 3 there exists a function that maps from $\Lambda \setminus \partial\Lambda$ to $\mathcal{S} \setminus \partial\mathcal{S}$. We need to show this function mapping is one-to-one and onto. First, given two distinct points $\mathbf{x}, \mathbf{y} \in \Lambda \setminus \partial\Lambda$, the function maps to $\mathbf{p}_{s,x}$, $\mathbf{p}_{s,y}$ respectively, both in $\mathcal{S} \setminus \partial\mathcal{S}$. If $\mathbf{p}_{s,x} = \mathbf{p}_{s,y}$, since they are both critical stabilizing controls (according to Prop. 3) meaning they determine the corresponding rate vectors $\mathbf{x} = \mathbf{x}(\mathbf{p}_{s,x})$, $\mathbf{y} = \mathbf{y}(\mathbf{p}_{s,y})$ according to (2.9), this gives $\mathbf{x} = \mathbf{y}$, which contradicts the assumption $\mathbf{x} \neq \mathbf{y}$ and hence this function is one-to-one. Second, for any point $\mathbf{p}_s \in \mathcal{S} \setminus \partial\mathcal{S}$ it defines a rate vector $\mathbf{x}(\mathbf{p}_s)$ according to (2.9), which by definition is in $\Lambda_{\text{eq}} = \Lambda$ and in fact is in $\Lambda \setminus \partial\Lambda$ (because of the bijection *i*)[16]). Recall $\mathbf{p}_s$ is automatically a critical stabilizing control for $\mathbf{x}(\mathbf{p}_s)$. That this function has to map $\mathbf{x}(\mathbf{p}_s)$ back to $\mathbf{p}_s$ is due to the fact that a rate vector from $\Lambda \setminus \partial\Lambda$ has exactly two critical stabilizing controls (one in $\mathcal{S} \setminus \partial\mathcal{S}$, the other in $[0,1]^n \setminus \mathcal{S}$), as shown at the end of the proof of Prop. 3. Therefore this function is onto. Thus we have shown the bijection *ii*). The proof of *iii*) is similar to that of *ii*) and is omitted. The proof of *iv*) follows from *ii*) and *iii*) due to transitivity. Finally *i*) and *ii*) together give *v*). □

Fig. 2.2 illustrates the three membership possibilities ($\mathbf{x} \in \Lambda \setminus \partial\Lambda$, $\mathbf{x} \in \partial\Lambda$, $\mathbf{x} \notin \Lambda$) and the corresponding polynomials $f(\delta, \mathbf{x})$ for the case $n = 2$. The case $n = 2$ is the only (known) value of n for which Λ can be expressed explicitly ([10],[5],[16]), i.e., $\Lambda = \Lambda_{\text{A}} = \left\{\mathbf{x} \in \mathbb{R}_+^2 : \sqrt{x_1} + \sqrt{x_2} \leq 1\right\}$.

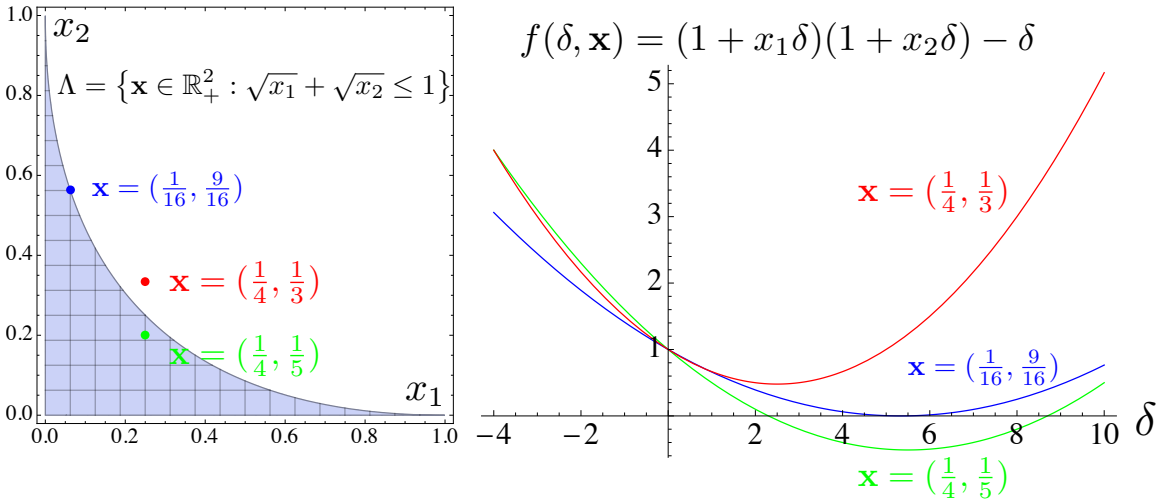

Figure 2.2: Polynomial root membership testing when $n = 2$. The green curve corresponds to the interior point $\mathbf{x} = (1/4, 1/5) \in \Lambda \setminus \partial\Lambda$, and has two positive roots: $\left((11 - \sqrt{41})/2, (11 + \sqrt{41})/2\right) \approx (2.2984, 8.7016)$; the blue curve corresponds to the boundary point $\mathbf{x} = (1/16, 9/16) \in \partial\Lambda$ and has a unique positive root $16/3 \approx 5.3333$; the red curve corresponds to a point $\mathbf{x} = (1/4, 1/3) \notin \Lambda$ and hence does not have any positive root.

2.3 Volume of Λ and an inner bound on Λ

We first give a closed-form expression for the volume of Λ. Unfortunately its computation is a formidable task.

Proposition 4. *The set* Λ *defined in* (2.1) *has volume*

$$\mathrm{vol}(\Lambda) \;=\; \sum_{\mathbf{k}\in\mathcal{K}_{2^n,n-2}} \binom{n-2}{\mathbf{k}} (-1)^{\sum_{i=1}^n \alpha(\mathbf{k})_i} \frac{\prod_{i=1}^n \alpha(\mathbf{k})_i!}{(n+1+\sum_{i=1}^n \alpha(\mathbf{k})_i)!}, \tag{2.14}$$

where $\binom{n-2}{\mathbf{k}}$ *is a multinomial coefficient and* $\mathcal{K}_{r,s} \equiv \left\{\mathbf{k}=(k_1,\ldots,k_r)\in\mathbb{Z}_+^r : \sum_{t=1}^r k_t = s\right\}$. *Furthermore,* $\alpha(\mathbf{k})_i = \sum_{t=1}^{2^n} \mathbf{V}_{i,t} k_t$ *for* $\mathbf{V}$ *the* $n\times 2^n$ *matrix whose columns are the* 2^n *possible length-n binary vectors.*

In coding theory, the matrix $\mathbf{V}$ is called a binary Hamming matrix.

Proof. Recall there is a bijection from $\mathcal{S}$ to Λ (Cor. 1). Let $\tilde{\mathbf{J}}(\mathbf{p}) \equiv \mathbf{J}(\mathbf{p})/\pi(\mathbf{p})$, where $\mathbf{J}(\mathbf{p})$ is the Jacobian of this mapping, namely the mapping $\mathbf{p}\mapsto\mathbf{x}$ given by (2.9) in Def. 2:

$$x_i(\mathbf{p}) = p_i \prod_{j\neq i}(1-p_j), \text{ for all } i\in[n],\ \mathbf{p}\in\mathcal{S}. \tag{2.15}$$

The fact that $\det(\alpha A) = \alpha^n \det A$ for any scalar α and any $n\times n$ matrix A yields $\det\mathbf{J}(\mathbf{p}) = \pi(\mathbf{p})^n \det\tilde{\mathbf{J}}(\mathbf{p})$. Abramson[9] showed that $\pi(\mathbf{p})^2 \det\tilde{\mathbf{J}}(\mathbf{p}) = 1-\mathbf{p}^\mathsf{T}\mathbf{1}$, which gives $\det\mathbf{J}(\mathbf{p}) = \pi(\mathbf{p})^{n-2}\left(1-\mathbf{p}^\mathsf{T}\mathbf{1}\right)$. Substituting this into the general expression for volume yields

$$\mathrm{vol}(\Lambda) = \int_{\mathcal{S}} \det\mathbf{J}(\mathbf{p})\mathrm{d}\mathbf{p} = \int_{\mathcal{S}} \prod_{i=1}^n (1-p_i)^{n-2}\left(1-\sum_{j=1}^n p_j\right)\mathrm{d}\mathbf{p}. \tag{2.16}$$

In order to get a better closed-form expression, we leverage results in Grundmann and Möller[19], in particular (2.3) on integration of certain functions over the solid standard unit simplex $\mathcal{S}$:

$$\int_{\mathcal{S}} \mathbf{p}^{\boldsymbol{\alpha}}\left(1-\sum_i p_i\right)^{\alpha_0}\mathrm{d}\mathbf{p} = \frac{\prod_{i=0}^n \alpha_i!}{(n+\sum_{i=0}^n \alpha_i)!}, \tag{2.17}$$

where $\mathbf{p} = (p_1, \ldots, p_n)$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$, and $\mathbf{p}^{\boldsymbol{\alpha}} = \prod_{i=1}^n p_i^{\alpha_i}$. To apply this general expression to our case (2.16), we want to put $\prod_{i=1}^n (1-p_i)^{n-2}$ into a weighted sum of terms of the form $\mathbf{p}^{\boldsymbol{\alpha}}$. The multi-binomial theorem states, for arbitrary n-vectors $\mathbf{a}, \mathbf{b}$, and positive n-vector $\mathbf{c}$:

$$\prod_{i=1}^n (a_i + y_i)^{c_i} = \sum_{k_1=0}^{c_1} \cdots \sum_{k_n=0}^{c_n} \binom{c_1}{k_1} a_1^{c_1-k_1} y_1^{k_1} \cdots \binom{c_n}{k_n} a_n^{c_n-k_n} y_n^{k_n}. \tag{2.18}$$

Specializing the above expression to the case $\mathbf{a} = \mathbf{1}$ and $\mathbf{c} = \mathbf{1}$ and arbitrary n-vector $\mathbf{y}$ yields:

$$\prod_{i=1}^n (1 + y_i)^1 = \sum_{k_1=0}^{1} \cdots \sum_{k_n=0}^{1} \binom{1}{k_1} 1^{1-k_1} y_1^{k_1} \cdots \binom{1}{k_n} 1^{1-k_n} y_1^{k_n} = \sum_{\mathbf{v} \in \{0,1\}^n} \binom{\mathbf{1}}{\mathbf{v}} \mathbf{1}^{\mathbf{1}-\mathbf{v}} \mathbf{y}^{\mathbf{v}} = \sum_{\mathbf{v} \in \{0,1\}^n} \mathbf{y}^{\mathbf{v}}, \tag{2.19}$$

where we employ the multi-index notation $\binom{\mathbf{a}}{\mathbf{b}} = \prod_{i=1}^n \binom{a_i}{b_i}$ and $\mathbf{a}^{\mathbf{b}} = \prod_{i=1}^n a_i^{b_i}$, for two n-vectors $\mathbf{a}, \mathbf{b}$. Consequently, for $\mathbf{y} = -\mathbf{p}$,

$$\prod_i (1-p_i)^{n-2} = \left(\sum_{\mathbf{v} \in \{0,1\}^n} (-\mathbf{p})^{\mathbf{v}} \right)^{n-2} = \left(\sum_{t=1}^{2^n} (-\mathbf{p})^{\mathbf{v}_t} \right)^{n-2} = \left(\sum_{t=1}^{2^n} \prod_{i=1}^{n} (-p_i)^{\mathbf{V}_{i,t}} \right)^{n-2}, \tag{2.20}$$

where $\mathbf{v}_t$ is the t^{th} column of $\mathbf{V}$. The multinomial theorem states, for arbitrary r-vector $\mathbf{y}$ and positive integer s,

$$(y_1 + \cdots + y_r)^s = \sum_{\mathbf{k} \in \mathcal{K}_{r,s}} \binom{s}{\mathbf{k}} \mathbf{y}^{\mathbf{k}}, \tag{2.21}$$

for $\mathcal{K}_{r,s}$ defined in the proposition. We apply the multinomial theorem to the RHS of (2.20) and get

$$\prod_i (1-p_i)^{n-2} = \sum_{\mathbf{k} \in \mathcal{K}_{2^n, n-2}} \binom{n-2}{k_1, \ldots, k_{2^n}} \prod_{t=1}^{2^n} \left(\prod_{i=1}^{n} (-p_i)^{\mathbf{V}_{i,t}} \right)^{k_t} = \sum_{\mathbf{k} \in \mathcal{K}_{2^n, n-2}} \binom{n-2}{k_1, \ldots, k_{2^n}} (-1)^{\sum_{i=1}^n \alpha_i} \prod_{i=1}^n p_i^{\alpha_i}, \tag{2.22}$$

for α_i defined in the proposition. Finally substitution of this expression of $\prod_i (1-p_i)^{n-2}$ into (2.16) and application of (2.17) with $\alpha_0 = 1$ yields the desired volume expression in (2.14). □

The number of summands in (2.14) is the number of multinomial coefficients. Equivalently, it is the number of ways to write $n-2$ as an ordered sum of 2^n non-negative integers, and is

given by $\binom{n-2+2^n-1}{n-2}$ (see e.g., Wilf[20] Ex. 3 in Chapter 2). Applying an easy lower bound on the binomial coefficient $\binom{n}{k} \geq \left(\frac{n}{k}\right)^k$, we have $\binom{n-2+2^n-1}{n-2} \geq \left(1+\frac{2^n-1}{n-2}\right)^{n-2}$, meaning it grows super-exponentially in n, and hence calculation of vol(Λ) using Prop. 4 requires substantial computation for even moderate n.

We now initiate our pursuit of non-parametric bounds on Λ, which is the focus of the next three sections. Recall it is already known that when $n = 2$, Λ equals a non-parametric set $\left\{\mathbf{x} \in \mathbb{R}^2_+ : \sqrt{x_1} + \sqrt{x_2} \leq 1\right\}$[5;10;16] for which membership testing is simple. Naturally one might wonder how the natural extension of this sum relates to Λ for higher values of n. This motivates the following definition of the "square root sum" set. The proposition below shows in general this set is only an inner bound on Λ. In the subsequent proof and elsewhere throughout the chapter, we use the fact that Λ is *coordinate convex*, meaning if $\mathbf{x} \in \Lambda$ then $\mathbf{x}' \in \Lambda$ for all $\mathbf{0} \leq \mathbf{x}' \leq \mathbf{x}$.

Definition 4.

$$\Lambda_{\text{srs}} \equiv \left\{\mathbf{x} \in \mathbb{R}^n_+ : \sum_{i=1}^n \sqrt{x_i} \leq 1\right\}. \tag{2.23}$$

Proposition 5 ("square root sum" inner bound)**.** *The set Λ_{srs} is an inner bound on Λ for $n \geq 2$.*

Proof. Fix a point $\mathbf{x}' \in \Lambda_{\text{srs}}$. Due to the coordinate convexity of Λ and Λ_{srs}, it suffices to produce a point $\mathbf{x} \in \Lambda$ so that $\mathbf{x} \geq \mathbf{x}'$. Set $\mathbf{p}$ with $p_i = \sqrt{x'_i}$ for each i and set $\mathbf{x} = \mathbf{x}(\mathbf{p})$ according to (2.9) in Def. 2. Clearly $\mathbf{x} \in \Lambda$. It remains to show $x_i \geq x'_i$ for each $i \in [n]$. Note $\mathbf{x}' \in \Lambda_{\text{srs}}$ ensures $\sum_{i=1}^n p_i \leq 1$. Define independent events $A_1, \ldots, A_n$ with $\mathbb{P}(A_i) = 1 - p_i$ for each $i \in [n]$. Denote the complement of event A_i by A_i^c. It follows that

$$1 - \mathbb{P}\left(\bigcup_{j\neq i} A_j^c\right) = \mathbb{P}\left(\bigcap_{j\neq i} A_j\right) = \prod_{j\neq i} \mathbb{P}\left(A_j\right) = \prod_{j\neq i}(1-p_j). \tag{2.24}$$

Then for any i, reversely applying (2.24) to x_i followed by the union bound and then the fact $\sum_{i=1}^n p_i \leq 1$, we have

$$x_i = p_i\left(1 - \mathbb{P}\left(\bigcup_{j\neq i} A_j^c\right)\right) \geq p_i\left(1 - \sum_{j\neq i}\mathbb{P}(A_j^c)\right) = p_i\left(1 - \sum_{j\neq i} p_j\right) \geq p_i^2 = x'_i. \tag{2.25}$$

□

Proposition 6. *The volume of the inner bound* Λ_{srs} *is*

$$\text{vol}(\Lambda_{\text{srs}}) = \frac{2^n}{(2n)!}. \tag{2.26}$$

Proof. Use the change of variable $y_i = \sqrt{x_i}$, $\forall i \in [n]$ so that the volume integration becomes

$$\begin{aligned}\text{vol}(\Lambda_{\text{srs}}) &= \int_{[0,1]^n} \mathbb{I}_{\sum_{i=1}^n \sqrt{x_i} \le 1}\ \mathrm{d}x_1 \cdots \mathrm{d}x_n = 2^n \int_{[0,1]^n} \mathbb{I}_{\sum_{i=1}^n y_i \le 1}\ y_1 \mathrm{d}y_1 \cdots y_n \mathrm{d}y_n \\ &= 2^n \int_0^1 y_n \int_0^{1-y_n} y_{n-1} \cdots \int_0^{1-y_n-\cdots-y_3} y_2 \int_0^{1-y_n-\cdots-y_2} y_1 \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_{n-1} \mathrm{d}y_n \end{aligned} \tag{2.27}$$

It will be useful to first compute an integral denoted $I(d,k) = \int_0^d y(d-y)^k \mathrm{d}y$ for $d \ge 0$, $k \in \mathbb{Z}_+$: instead of expanding the integrand using binomial theorem and then handling some alternating sum, we proceed as follows:

$$\begin{aligned} -I(d,k) + \int_0^d d(d-y)^k \mathrm{d}y &= \int_0^d (d-y)^{k+1} \mathrm{d}y = \frac{1}{k+2} d^{k+2} \\ \Rightarrow I(d,k) = \int_0^d d(d-y)^k \mathrm{d}y - \frac{1}{k+2} d^{k+2} &= d\frac{1}{k+1} d^{k+1} - \frac{1}{k+2} d^{k+2} = \frac{1}{(k+1)(k+2)} d^{k+2} \end{aligned} \tag{2.28}$$

For $j \in [n]$ define $k_j = 2(j-1)$ and for $j \in [n-1]$ define $d_j = 1 - y_n - \cdots - y_{j+1}$, and $d_n = 1$. Observe the recurrences $k_j + 2 = k_{j+1}$, $d_j = d_{j+1} - y_{j+1}$. Specializing (2.28) with parameters d_j, k_j and dummy integrating variable y_j, we have

$$I(d_j, k_j) = \int_0^{d_j} y_j (d_j - y_j)^{k_j} \mathrm{d}y_j = \frac{1}{(k_j+1)(k_j+2)} d_j^{k_j+2},\ \forall j \in [n]. \tag{2.29}$$

Now we are ready to resume the computation of $\text{vol}(\Lambda_{\text{srs}})$ in (2.27). Using our new notation, we have:

$$\text{vol}(\Lambda_{\text{srs}}) \quad = \quad 2^n \int_0^{d_n} y_n \int_0^{d_{n-1}} y_{n-1} \cdots \int_0^{d_2} y_2 \int_0^{d_1} y_1 (d_1 - y_1)^{k_1} \mathrm{d}y_1 \mathrm{d}y_2 \cdots \mathrm{d}y_{n-1} \mathrm{d}y_n. \tag{2.30}$$

We can then repeatedly apply (2.29) with $j \in [n]$. To see this, observe after the j^{th} innermost integration, the new innermost integration is

$$\int_0^{d_{j+1}} \prod_{s=1}^{j} \frac{1}{(k_s+1)(k_s+2)} y_{j+1} d_j^{k_j+2} \mathrm{d}y_{j+1} = \prod_{s=1}^{j} \frac{1}{(k_s+1)(k_s+2)} \int_0^{d_{j+1}} y_{j+1} \left(d_{j+1} - y_{j+1}\right)^{k_{j+1}} \mathrm{d}y_{j+1}, \tag{2.31}$$

which is $\prod_{s=1}^{j} \frac{1}{(k_s+1)(k_s+2)} I(d_{j+1}, k_{j+1})$.

Therefore, after the $j = (n-1)^{\text{st}}$ innermost integration, we have

$$\begin{aligned} \mathrm{vol}(\Lambda_{\text{srs}}) &= 2^n \prod_{s=1}^{n-1} \frac{1}{(k_s+1)(k_s+2)} \int_0^{d_n} y_n d_{n-1}^{k_{n-1}+2} \mathrm{d}y_n = 2^n \prod_{s=1}^{n-1} \frac{1}{(k_s+1)(k_s+2)} \int_0^{d_n} y_n \left(d_n - y_n\right)^{k_n} \mathrm{d}y_n \\ &= 2^n \prod_{s=1}^{n-1} \frac{1}{(k_s+1)(k_s+2)} I(d_n, k_n) = 2^n \prod_{s=1}^{n} \frac{1}{(k_s+1)(k_s+2)} d_n^{k_n+2} = 2^n \prod_{s=1}^{n} \frac{1}{(2s-1)(2s)} \\ &= \frac{2^n}{(2n)!}. \end{aligned} \tag{2.32}$$

□

Although simple, as has been seen in Fig. 2.1 (in §2.1), Λ_{srs} is a poor inner bound. In the following three sections we present various inner and outer bounds on Λ based on polyhedra (§2.4), spheres (§2.5), and ellipsoids (§2.6).

2.4 Polyhedral inner and outer bounds on Λ

In this section we form inner and outer bounds on Λ using polyhedra. The inner bound is formed using a single hyperplane, i.e., a generalized simplex, while the outer bound is formed using the intersection of a collection of $n+1$ hyperplanes in $\mathbb{R}^n_+$.

Definition 5.

$$\Lambda_{\text{pi}}(\mathbf{p}) \equiv \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \ (\mathbf{1}-\mathbf{p})^{\mathsf{T}} \mathbf{x} \le \prod_i (1-p_i) \right\}, \ \textit{where } \mathbf{p} \in \partial\mathcal{S}. \tag{2.33}$$

Geometrically, the set $\Lambda_{\text{pi}}(\mathbf{p})$ is a generalized simplex bounded by the n coordinate hyperplanes and the hyperplane with normal vector $\mathbf{1}-\mathbf{p}$. All such hyperplanes are tangent to $\partial\Lambda$ with a

tangency point at $\mathbf{x}(\mathbf{p})$ in (2.9) in Def. 2. The following proposition asserts for each given $\mathbf{p}$ this set is an inner bound on Λ, indicates the $\mathbf{p}^*$ that achieves the largest volume bound over this family of inner bounds, and also computes the corresponding volume.

Proposition 7 (polyhedral inner bound)**.** *For each $\mathbf{p} \in \partial\mathcal{S}$, the set $\Lambda_{\text{pi}}(\mathbf{p})$ is an inner bound on Λ for $n \geq 2$. Among these, the tightest is given when $\mathbf{p} = \mathbf{p}^* = \frac{1}{n}\mathbf{1}$, namely,*

$$\Lambda^*_{\text{pi}} = \Lambda_{\text{pi}}(\mathbf{p}^*) = \left\{ \mathbf{x} \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i \leq \left(1 - \frac{1}{n}\right)^{n-1} \right\}. \tag{2.34}$$

and the corresponding volume of this set is $\text{vol}(\Lambda^*_{\text{pi}}) = \frac{1}{n!}\left(1 - \frac{1}{n}\right)^{n(n-1)}$.

Proof. Recall Post[17] established that the complement of Λ in $\mathbb{R}^n_+$ is convex, and gave the tangent hyperplane at a point $\mathbf{x}(\mathbf{p})$ on $\partial\Lambda$: $\left\{\mathbf{x} : (\mathbf{1} - \mathbf{p})^\mathsf{T}\mathbf{x} = \prod_i (1 - p_i)\right\}$, where $\mathbf{p} \in \partial\mathcal{S}$ is the unique control associated with a point $\mathbf{x} \in \partial\Lambda$. Since this hyperplane is a supporting hyperplane, this open convex set $\Lambda^c \cap \mathbb{R}^n_+$ lies entirely on one "side", i.e., the open halfspace $\left\{\mathbf{x} : (\mathbf{1} - \mathbf{p})^\mathsf{T}\mathbf{x} > \prod_i (1 - p_i)\right\}$, of the hyperplane. This means points on the other side of this hyperplane are not in $\Lambda^c \cap \mathbb{R}^n_+$, and hence are in Λ, i.e., $\Lambda_{\text{pi}}(\mathbf{p}) \subseteq \Lambda$.

Now notice $\Lambda_{\text{pi}}(\mathbf{p})$ is a generalized simplex, and its volume is given by ([21]):

$$\text{vol}(\Lambda_{\text{pi}}(\mathbf{p})) = \frac{1}{n!}\prod_{i=1}^n \prod_{j \neq i} (1 - p_j) = \frac{1}{n!}\left(\prod_{i=1}^n (1 - p_i)\right)^{n-1}. \tag{2.35}$$

It is easily shown that the function $\prod_{i=1}^n (1 - p_i)$ is maximized over $\mathbf{p} \in \partial\mathcal{S}$ at $\mathbf{p} = \frac{1}{n}\mathbf{1}$, and hence the best Λ_{pi} (in terms of achieving the largest volume) is given by (2.34). □

Remark 1. *Using lower and upper bounds on the factorial[22], one can show* $\text{vol}(\Lambda^*_{\text{pi}}) \geq \text{vol}(\Lambda_{\text{srs}})$ *for all $n \geq 3$.*

Next we construct a polyhedral outer bound. If we restrict ourselves to only using a single half-space, the best choice is the standard simplex, $\mathcal{S}$, which is a very loose outer bound. Consequently, we consider a specific construction using $2n + 1$ hyperplanes. The convex polytope given below is a subset of $\mathcal{S}$ (and in fact a subset of $\overline{\mathcal{S} \setminus \Lambda}$), has $\partial\mathcal{S}$ as a facet, and an additional n facets each defined

by a hyperplane, $(\mathcal{H}_1^\alpha, \ldots, \mathcal{H}_n^\alpha)$, where $\mathcal{H}_i^\alpha$ is the hyperplane passing through $\mathbf{e}_i, \mathbf{m}$, and $\alpha\mathbf{e}_j$ for all $j \neq i$, for α given below.

Definition 6. *The halfspace representation of the convex polytope* P *in* $\mathbb{R}^n$ *consists of the following halfspaces:*

$$\begin{aligned}
\mathcal{H}_i^{\alpha+} &\equiv \left\{ \mathbf{x} \in \mathbb{R}^n : \; x_i + \frac{1}{\alpha(n)} \sum_{j \neq i} x_j \geq 1 \right\}, \; i \in [n], \\
\mathcal{H}_i^{\mathrm{c}+} &\equiv \left\{ \mathbf{x} \in \mathbb{R}^n : \; x_i \geq 0 \right\}, \; i \in [n], \\
\mathcal{H}^{\partial\mathcal{S}-} &\equiv \left\{ \mathbf{x} \in \mathbb{R}^n : \; \sum_{i=1}^n x_i \leq 1 \right\},
\end{aligned} \tag{2.36}$$

where

$$\alpha(n) \equiv \frac{n-1}{\frac{n}{\left(1-\frac{1}{n}\right)^{n-1}} - 1} = \frac{n-1}{\frac{1}{m(n)} - 1} \tag{2.37}$$

and the superscript $^+$ *indicates an "upward" halfspace and* $^-$ *indicates a "downward" halfspace. More compactly,*

$$\mathsf{P} \equiv \left\{ \mathbf{x} \in \mathbb{R}^n : \; \mathbf{x} \in \bigcap_{i \in [n]} \mathcal{H}_i^{\alpha+} \bigcap_{i \in [n]} \mathcal{H}_i^{\mathrm{c}+} \cap \mathcal{H}^{\partial\mathcal{S}-} \right\}. \tag{2.38}$$

Furthermore, the corresponding hyperplane is denoted by dropping these superscripts, meaning the inequality in the definition holds with equality. For example, $\mathcal{H}_i^{\mathrm{c}}$ *denotes the coordinate hyperplane* $\{\mathbf{x} \in \mathbb{R}^n : x_i = 0\}$.

Proposition 8 (polyhedral outer bound)**.** *The convex polytope* P *defined above induces an outer bound on* Λ*. More precisely,* $\Lambda \subseteq \Lambda_{\mathrm{po}} \equiv \mathcal{S} \setminus \mathsf{P}$.

To prove the correctness of this bound it will be essential to establish the monotonicity of $\alpha(n)$ (2.37).

Lemma 1. *The function* $\alpha(n)$ (2.37) *is monotone increasing for* $n \geq 2$*. In particular,* $\alpha(2) = 1/3$*,* $\alpha(3) = 8/23$*,* $\alpha(\infty) = 1/\mathrm{e}$.

Proof. The derivative of $\alpha(n)$ (2.37) is

$$\frac{\mathrm{d}\alpha(n)}{\mathrm{d}n} = \frac{(n-1)^2 \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}}{\left(\left(1-\frac{1}{n}\right)^n - (n-1)\right)^2}\left(1 - \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} + (n-1)\log\left(1-\frac{1}{n}\right)\right), \tag{2.39}$$

for which the sign is determined by the sign of $g(n) \equiv 1 - \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} + (n-1)\log\left(1-\frac{1}{n}\right)$. To show the positivity of $g(n)$ for all $n \geq 2$, first observe $\lim_{n\to\infty} g(n) = 0$. It therefore suffices to show $g(n)$ is itself monotone decreasing in n, which is shown below:

$$\begin{aligned} n\frac{\mathrm{d}g(n)}{\mathrm{d}n} &= 1 + n\left(1 - \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}\right)\log\left(1-\frac{1}{n}\right) \\ &\overset{(a)}{\leq} 1 + n\left(1 - \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}\right)\left(-\frac{1}{n} - \frac{1}{2n^2} - \frac{1}{3n^3}\right) \\ &\overset{(b)}{\leq} 1 + n\left(1 - \frac{1}{2n}\right)\left(-\frac{1}{n} - \frac{1}{2n^2} - \frac{1}{3n^3}\right) = -\frac{n-2}{12n^3} \leq 0, \end{aligned} \tag{2.40}$$

where we use in (a) the inequality $\log(1+x) \leq x - \frac{x^2}{2} + \frac{x^3}{3}$ for all $x \in (-1, 0]$ and (b) the property that $\left(1-\frac{1}{n}\right)^{n-1}$ is monotone decreasing in n from 1/2 (when $n = 2$) to 1/e (when $n = \infty$). □

Proof. (of Prop. 8) Our approach is to show all the vertices of the convex polytope P are in $\overline{\Lambda^c}$, it then follows from the convexity of $\overline{\Lambda^c} \cap \mathbb{R}^n_+$ that $\mathsf{P} \subseteq \overline{\Lambda^c} \cap \mathbb{R}^n_+$ which implies $\Lambda_{\mathrm{po}} \equiv \mathcal{S} \setminus \mathsf{P} \supseteq \Lambda$.

To find a vertex, we first choose n out of the $2n+1$ hyperplanes defining P. If there is a solution to this linear system which is a single point that *also* obeys the remaining $n+1$ halfspace constraints, then this solution is a valid vertex (indicated below as underlined cases). Furthermore, since our primary goal is to show all the vertices are in $\overline{\Lambda^c}$, rather than to list all the vertices, for simplicity we only consider the scenario where those n selected hyperplanes do not include $\mathcal{H}^{\partial\mathcal{S}}$. This is justified since the intersection of $\mathcal{H}^{\partial\mathcal{S}}$ and n halfspaces $\mathcal{H}_i^{\mathrm{c}+}$, namely $\partial\mathcal{S}$, lies completely in $\overline{\Lambda^c}$, so if there exists a valid vertex on $\mathcal{H}^{\partial\mathcal{S}}$ it is guaranteed to be in $\overline{\Lambda^c}$.

Consequently, we choose a set of hyperplanes from $\{\mathcal{H}_i^{\alpha}\}_{i=1}^n$ (denoted S) and a set of hyperplanes from $\{\mathcal{H}_i^{\mathrm{c}}\}_{i=1}^n$ (denoted T) so that their cardinalities $|S|$ and $|T|$ sum to n. We also assume we choose the first $|S|$-indexed hyperplanes from $\{\mathcal{H}_i^{\alpha}\}_{i=1}^n$; this holds with no loss of generality as we may always permute the indices of the hyperplanes, and the polytope P is symmetric with respect

to such permutations. For notational convenience define $\mathcal{I}_S$ and $\mathcal{I}_T$ as the set of indices appearing as subscripts of the elements in the set S and T respectively. For example, if $S = \{\mathcal{H}_1^\alpha, \mathcal{H}_2^\alpha\}$, then $\mathcal{I}_S = \{1, 2\}$; if $T = \{\mathcal{H}_1^{\mathrm{c}}\}$, then $\mathcal{I}_T = \{1\}$. Recall, $m(k) = \frac{1}{k}\left(1 - \frac{1}{k}\right)^{k-1}$ for $k \in [n]$ stands for the coordinate of the all-rates-equal point on $\partial\Lambda$ in a k-dimensional space.

We discuss cases based on the pair $(|\mathcal{I}_S \cap \mathcal{I}_T|, |S|)$

- case 1: $|\mathcal{I}_S \cap \mathcal{I}_T| = 0$, $|S| = n$. Namely we choose all n $\mathcal{H}_i^\alpha$'s, to which the only solution is all-rates-equal point $\mathbf{m} = m\mathbf{1}$, which is in $\overline{\Lambda^c}$.

- case 2: $|\mathcal{I}_S \cap \mathcal{I}_T| = 0$, $|S| = k$ for $1 \le k < n$. Namely $S = \{\mathcal{H}_1^\alpha, \ldots, \mathcal{H}_k^\alpha\}$, $T = \left\{\mathcal{H}_{k+1}^{\mathrm{c}}, \ldots, \mathcal{H}_n^{\mathrm{c}}\right\}$. The only solution can be shown to be $\mathbf{x} = \left(1 + \frac{k-1}{\alpha(n)}\right)^{-1} \sum_{j=1}^{k} \mathbf{e}_j$. To verify this point $\mathbf{x}$ is in P, we first verify it satisfies the halfspace constraint $\mathcal{H}_{k+1}^{\alpha+}$, i.e., $x_{k+1} + \frac{1}{\alpha(n)} \sum_{j \neq k+1} x_j \ge 1$, which applied to this point becomes $\alpha(n) \le 1$. Similarly $\mathbf{x}$ also satisfies the halfspace constraints associated with $\mathcal{H}_{k+2}^\alpha$, ..., $\mathcal{H}_n^\alpha$. Next, for $\mathbf{x}$ to satisfy the halfspace constraint $\mathcal{H}^{\partial\mathcal{S}-}$, we again only need $\alpha(n) \le 1$. Finally the nonnegativity constraint for each coordinate axis $\mathcal{H}_i^{\mathrm{c}+}$ is satisfied, so this solution is a valid vertex. We now need to show $\mathbf{x} \in \overline{\Lambda^c}$. Observe this vertex only has k non-zero components so we need to check $\overline{\Lambda^c}$ in the corresponding k-dimensional space; furthermore, all the non-zero components of $\mathbf{x}$ are identical meaning $\mathbf{x}$ lies along the all-rates-equal ray in this k-dimensional space so we only need to show $\mathbf{x}$ extends beyond the corresponding all-rates-equal point $\mathbf{m} = m(k)\mathbf{1}$ for $\mathbf{1}$ a k-vector of all 1's. Applying Lem. 1, we have $\left(1 + \frac{k-1}{\alpha(n)}\right)^{-1} \ge \left(1 + \frac{k-1}{\alpha(k)}\right)^{-1} = m(k)$. Thus we've shown this case does produce a valid vertex in $\overline{\Lambda^c}$.

- case 3: $|\mathcal{I}_S \cap \mathcal{I}_T| = 0$, $|S| = 0$. Namely we choose all n coordinate hyperplane $\mathcal{H}_i^c$'s. The only solution is the origin $\mathbf{o}$, which is not in P, hence this is an invalid vertex.

- case 4: $|\mathcal{I}_S \cap \mathcal{I}_T| = 1$, $|S| = k$ for $1 \le k < n$. In this case, in order to further satisfy $|S| + |T| = n$, there must exist some index $k' \ge k+1$ such that $\mathcal{H}_{k'}^c \notin T$. In fact if we assume $\mathcal{H}_1^{\mathrm{c}} \in T$, this determines $T = \{\mathcal{H}_1^{\mathrm{c}}, \mathcal{H}_{k+1}^{\mathrm{c}}, \ldots, \mathcal{H}_n^{\mathrm{c}}\} \setminus \{\mathcal{H}_{k'}^c\}$. The solution can be shown to be $\mathbf{x} = \alpha\mathbf{e}_{k'}$, which is not in P, and hence is not a valid vertex. Note this conclusion does not

depend on our choice of $\mathcal{H}_1^{\mathrm{c}}$ to be included in T.

- case 5: $|\mathcal{I}_S \cap \mathcal{I}_T| > 1$, $|S| = k$ for $1 < k < n-1$. In this case, in order to further satisfy $|S| + |T| = n$, there must exist $l = |\mathcal{I}_S \cap \mathcal{I}_T|$ indices such that the corresponding coordinate hyperplanes are not in T. Attempt to solve this system shows this is an underdetermined system because the solution is given by a hyperplane instead of a point. Furthermore, if we want to ensure the solution is in P, we find there is no consistent solution. As $S = \{\mathcal{H}_1^{\alpha}, \ldots, \mathcal{H}_k^{\alpha}\}$, suppose T does not include, say, $\mathcal{H}_{k+1}^{\mathrm{c}}, \ldots, \mathcal{H}_{k+l}^{\mathrm{c}}$ (as well as $\mathcal{H}_{l+1}^{\mathrm{c}}, \ldots, \mathcal{H}_k^{\mathrm{c}}$), so $T = \{\mathcal{H}_1^{\mathrm{c}}, \ldots, \mathcal{H}_l^{\mathrm{c}}, \mathcal{H}_{k+l+1}^{\mathrm{c}}, \ldots, \mathcal{H}_n^{\mathrm{c}}\}$. Solving these n equations gives an l-dimensional hyperplane: $\{\mathbf{x} : x_{k+1} + \cdots + x_{k+l} = \alpha\}$. Satisfying the halfspace constraint $\mathcal{H}_{k+1}^{\alpha+}$ as well as each nonnegativity component constraints $\mathcal{H}_i^{\mathrm{c}+}$ requires $x_{k+1} = 0$. Similarly, due to each other halfspace constraint $\mathcal{H}_{k+2}^{\alpha+}, \ldots, \mathcal{H}_{k+l}^{\alpha+}$, each other component $x_{k+2}, \ldots, x_{k+l}$ would also need to be set to zero. These together lead to no valid (vertex) solution.

To summarize, each valid vertex solution we have found is such that: *i*) its non-zero components are all equal, and *ii*) this non-zero component value is no smaller than the coordinate of all-rates-equal point in the corresponding possibly reduced-dimensional space, which means all those vertices are in $\overline{\Lambda^c}$. More precisely, each vertex extends beyond (or coincides with) the corresponding all-rates-equal point (which lies on the boundary of Λ), and can be written as (up to permutation of the indices) $\mathbf{x} = \left(1 + \frac{k-1}{\alpha(n)}\right)^{-1} \sum_{j=1}^{k} \mathbf{e}_j$ for $k \in [n]$. In particular, when $k = 1$, $\mathbf{x} = \mathbf{e}_1$; when $k = n$, $\mathbf{x} = \mathbf{m}$. □

Remark 2. *One can perform a similar analysis considering the scenario where $\mathcal{H}^{\partial\mathcal{S}}$ is selected. The only valid vertex solutions consist of just $\mathbf{e}_i$'s for all $i \in [n]$.*

We give the vertex representation of the convex polytope P when $n = 2$ and $n = 3$ in the following example. The bounds Λ_{pi}^* and Λ_{po} together with $\partial\Lambda$ are illustrated in Fig. 2.3.

Example 1. *When $n = 2$, since $\alpha(2) = 1/3$, the vertices of P are $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{m}\}$, here $\mathbf{m} = m(2)\mathbf{1}$ for $m(2) = 1/4$. When $n = 3$, since $\alpha(3) = 8/23$, the vertices of P are $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, (8/31)(\mathbf{e}_1 + \mathbf{e}_2), (8/31)(\mathbf{e}_1 + \mathbf{e}_3), (8/31)(\mathbf{e}_2 + \mathbf{e}_3), \mathbf{m}\}$, here $\mathbf{m} = m(3)\mathbf{1}$ for $m(3) = 4/27$. Those vertices are also*

shown in Fig. 2.3. Note $8/31 > 1/4$ *thus each of the three green points in the right subfigure extends beyond the all-rates-equal point on the corresponding* 2*-dimensional plane namely the orange point in the left subfigure.*

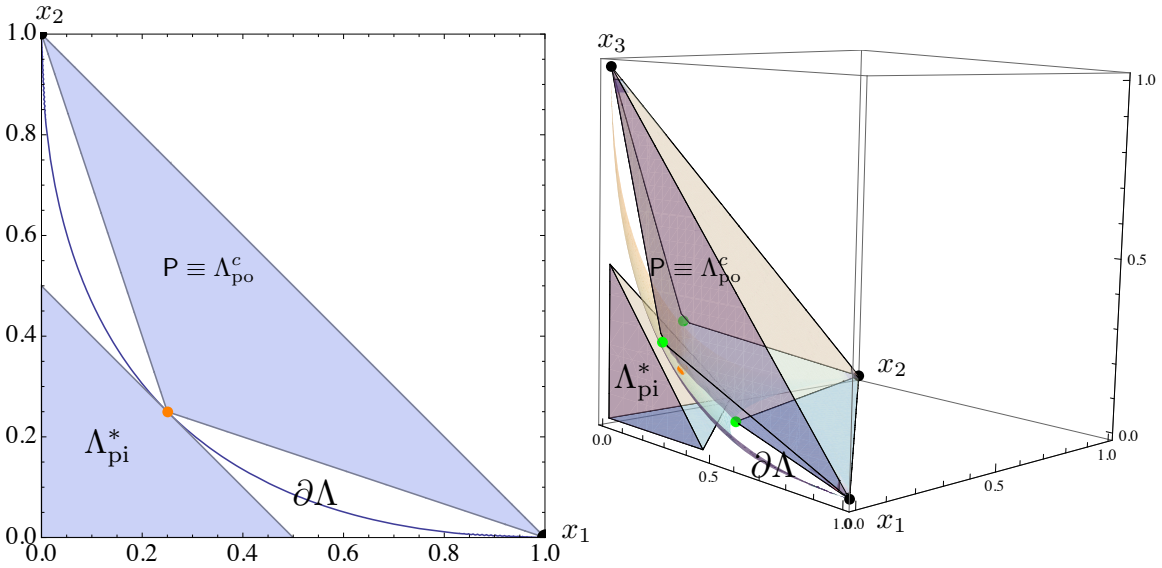

Figure 2.3: Polyhedral bounds for $n = 2$ (left) and 3 (right). The inner bound and the complement (w.r.t. $\mathcal{S}$) of the outer bound (namely P, recall $\Lambda_{\rm po} \equiv \mathcal{S} \setminus \mathsf{P}$) are shown in solid, and in between sits $\partial\Lambda$. Also shown are the vertices of the polytope used in $\Lambda_{\rm po}$: black points are $\mathbf{e}_i$'s, the orange point is the all-rates-equal point $\mathbf{m}$, the green points on the right have all non-zero components equal $8/31$.

2.5 Spherical inner and outer bounds on Λ

In this section we consider bounds induced by spheres. More specifically we want $\mathcal{S} \setminus \mathcal{B}(\mathbf{c}, r)$ to be included in (for inner bounding) or to include (for outer bounding) Λ, where $\mathcal{B}(\mathbf{c}, r)$ denotes the open ball in $\mathbb{R}^n$ with center $\mathbf{c}$ and radius r, and $\partial\mathcal{B}(\mathbf{c}, r)$ is its boundary. By symmetry, we restrict our attention to balls with centers on the all-rates-equal ray, i.e., $\mathbf{c} = c\mathbf{1}$ for some $c > 0$. In the following, Prop. 9 establishes a family of inner bounds induced by balls centered at $\mathbf{c} = c\mathbf{1}$ with radius $r(c) = d(\mathbf{c}, \mathbf{m})$ (i.e., $\mathbf{m} \in \partial\mathcal{B}(\mathbf{c}, r)$), and among them the best one (in the sense of giving the best approximation of the volume of Λ) is obtained by $c = (1 - nm^2)/(2(1 - nm))$, which is indeed the minimum c in order to *possibly* produce a valid spherical inner bound in this family. A parallel result, Prop. 10, establishes a family of outer bounds induced by balls centered at $\mathbf{c} = c\mathbf{1}$ with radius defined as $r(c) = d(\mathbf{c}, \mathbf{e}_i)$ (i.e., $\mathbf{e}_i \in \partial\mathcal{B}(\mathbf{c}, r)$ for $i \in [n]$), and among them the best one is given when

$c = 1$, which is also the minimum c in order to produce a valid spherical outer bound in this family.

Definition 7. *$\Lambda_{\rm si}(c) \equiv \mathcal{S} \setminus \mathcal{B}(\mathbf{c}, r_{\rm in}(c))$, where the center of the ball is $\mathbf{c} = c\mathbf{1}$ for all $c \geq c^*_{\rm in} \equiv (1 - nm^2)/(2(1 - nm))$, and its radius $r_{\rm in}(c) \equiv d(\mathbf{c}, \mathbf{m}) = \sqrt{n}(c - m)$.*

Proposition 9 (spherical inner bound)**.** *For each $c \geq c^*_{\rm in}$, the set $\Lambda_{\rm si}(c)$ is an inner bound on Λ for $n \geq 2$. Among these, the tightest is given when $c = c^*_{\rm in}$:*

$$\Lambda^*_{\rm si} = \Lambda_{\rm si}(c^*_{\rm in}) = \left\{\mathbf{x} \in \mathcal{S} : \|\mathbf{x} - c^*_{\rm in}\mathbf{1}\| \geq \sqrt{n}\,(c^*_{\rm in} - m)\right\}. \tag{2.41}$$

Proof. Here is an overview of the proof. First, we observe the correctness of the spherical inner bound with some c implies the correctness of an inferior bound with a larger c (Lem. 2), so we only need to show the correctness of the bound with the minimum c namely $c^*_{\rm in}$. Second, by inspecting the Karush-Kuhn-Tucker (KKT) conditions, we argue a potential local extremizer can have at most two distinct non-zero component values, and also obtain a condition the components of this extremizer must satisfy ((2.54)). Third, we address the case when a potential extremizer does not have zero component and has exactly two distinct non-zero component values, and we show such a point can be safely ruled out for the optimization problem set up in Step 2. Fourth, we consider the case when a potential extremizer has zero component(s), and show this point can be removed too (unless it reduces to $\mathbf{e}_i$). Finally, it is clear we only need to evaluate the objective function at $\mathbf{m}$ and $\mathbf{e}_i$.

Step 1: correctness of the bound with small c implies correctness and inferiority of the bound with larger c. Lem. 2 below establishes that $c_2 \geq c_1 \geq m$ implies $\mathcal{B}(\mathbf{c}_2, d(\mathbf{c}_2, \mathbf{m})) \supseteq \mathcal{B}(\mathbf{c}_1, d(\mathbf{c}_1, \mathbf{m}))$, i.e., the balls in this family are nested in c. Since $\Lambda_{\rm si}(c) = \mathcal{S} \setminus \mathcal{B}(\mathbf{c}, r_{\rm in}(c))$, it follows that $c_2 \geq c_1 \geq m$ implies $\Lambda_{\rm si}(c_1) \supseteq \Lambda_{\rm si}(c_2)$, i.e., the induced bounds are likewise nested, and thus the optimal (largest) bound in this family is obtained by the smallest c in the family. Because of this, we need only establish that $\Lambda_{\rm si}(c) \subseteq \Lambda$ for this smallest c in the family. To establish $c^*_{\rm in}$ is the minimum c, it suffices to verify the following:

$$d(\mathbf{c}, \mathbf{m}) <, =, > d(\mathbf{c}, \mathbf{e}_i) \text{ if and only if } c <, =, > c^*_{\rm in} \text{ respectively.} \tag{2.42}$$

This is straightforward to establish, and the proof is omitted. Assuming (2.42) to be true, it follows that if $c < c^*_{\text{in}}$ then each $\mathbf{e}_i$ will not be included in the closed ball $\overline{\mathcal{B}}(\mathbf{c}, r_{\text{in}})$, implying the induced bound $\mathcal{S} \setminus \mathcal{B}(\mathbf{c}, r_{\text{in}})$ is invalid (since each $\mathbf{e}_i \in \Lambda$).

Lemma 2. *$\mathcal{B}(\mathbf{c}_2, d(\mathbf{c}_2, \mathbf{m})) \supseteq \mathcal{B}(\mathbf{c}_1, d(\mathbf{c}_1, \mathbf{m}))$ for $c_2 \geq c_1 \geq m$.*

Proof of Lem. 2: For simplicity we shift the origin of coordinate system along the all-rates-equal ray $\mathbf{1}$ so that it overlaps with $\mathbf{m}$ in the original system. In the new system we have $c'_2 = c_2 - m$, $c'_1 = c_1 - m$, and $m' = 0$, and we need to show $\mathcal{B}(\mathbf{c}'_2, d(\mathbf{c}'_2, \mathbf{o}')) \supseteq \mathcal{B}(\mathbf{c}'_1, d(\mathbf{c}'_1, \mathbf{o}'))$ for $c'_2 \geq c'_1 \geq 0$, where $\mathbf{o}'$ denotes the origin of the new system. Observe $d(\mathbf{c}'_j, \mathbf{o}') = \sqrt{n}c'_j$ for $j = 1, 2$. So we need to verify for all $\mathbf{x}$ satisfying $\sum_i (x_i - c'_1)^2 \leq nc'^2_1$, it holds that $\sum_i (x_i - c'_2)^2 \leq nc'^2_2$. Towards this, we write

$$\sum_i (x_i - c'_2)^2 = \sum_i (x_i - c'_1)^2 + n(c'_2 - c'_1)^2 - 2(c'_2 - c'_1) \sum_i (x_i - c'_1) \leq nc'^2_1 + n(c'_2 - c'_1)^2 - 2(c'_2 - c'_1) \sum_i (x_i - c'_1). \tag{2.43}$$

So it suffices to show the RHS is no larger than nc'^2_2, which is equivalent to showing $\sum_i x_i \geq 0$. We claim this is true, because the hyperplane $\{\mathbf{x} \in \mathbb{R}^n : \sum_i x_i = 0\}$ is tangent with $\mathcal{B}(\mathbf{c}'_1, d(\mathbf{c}'_1, \mathbf{o}'))$ at $\mathbf{o}'$, and in fact it is a supporting hyperplane of the convex body $\mathcal{B}(\mathbf{c}'_1, d(\mathbf{c}'_1, \mathbf{o}'))$. □

Steps 2, 3, and 4 are actually valid for $\Lambda_{\text{si}}(c)$ for all c, not just $c = c^*_{\text{in}}$, and so we consider an arbitrary $\Lambda_{\text{si}}(c)$ in what follows.

Step 2: properties of a potential extremizer for Λ_{si}. By Prop. 15 in §2.6.1, it suffices to establish $\partial\Lambda \subseteq \overline{\mathcal{B}}(\mathbf{c}, r_{\text{in}})$, i.e., given any point $\mathbf{x} \in \partial\Lambda$, its distance to the center of the sphere is no larger than the sphere's radius:

$$\max_{\mathbf{x} \in \partial\Lambda} d(\mathbf{c}, \mathbf{x})^2 \leq r^2_{\text{in}}. \tag{2.44}$$

Recall from Cor. 1 (§2.2) the bijection between $\partial\Lambda$ and $\partial\mathcal{S}$, and write $\mathbf{x}(\mathbf{p})$ to denote the unique $\mathbf{x} \in \partial\Lambda$ associated with each $\mathbf{p} \in \partial\mathcal{S}$. Under this bijection, the LHS of (2.44) becomes

$$\max_{\mathbf{p} \in \partial\mathcal{S}} f(\mathbf{p}) \equiv d(\mathbf{c}, \mathbf{x}(\mathbf{p}))^2 = \sum_{i=1}^{n} \left(c - p_i \prod_{j \neq i}^{n} (1 - p_j) \right)^2. \tag{2.45}$$

Introducing Lagrange multipliers μ, $(\lambda_i, i \in [n])$ for the equality constraint and n inequality constraints in $\partial\mathcal{S}$, respectively, the Lagrangian of this maximization problem becomes:

$$\mathcal{L}(\mathbf{p}, \mu, \boldsymbol{\lambda}) = f(\mathbf{p}) + \mu \left(\sum_{i=1}^{n} p_i - 1 \right) + \sum_{i=1}^{n} \lambda_i \left(-p_i \right). \tag{2.46}$$

The first-order Karush-Kuhn-Tucker (KKT) necessary conditions for a local maximizer are:

$$\text{stationarity} \qquad \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p_i} = 0, \ i \in [n] \tag{2.47}$$

$$\text{primal feasibility} \qquad \sum_{i=1}^{n} p_i - 1 = 0, \ -p_i \leq 0, \ i \in [n] \tag{2.48}$$

$$\text{dual feasibility} \qquad \lambda_i \leq 0, \ i \in [n] \tag{2.49}$$

$$\text{complementary slackness} \qquad \lambda_i \left(-p_i \right) = 0, \ i \in [n]. \tag{2.50}$$

Note the regularity condition LICQ (linear independence constraint qualification) is satisfied.

Observe $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p_i} = \frac{\mathrm{d}f}{\mathrm{d}p_i} + \mu - \lambda_i$. Therefore, if a potential local maximizer $\mathbf{p}$ has two distinct non-zero components $0 < p_k < p_l$ then, by complementary slackness, stationarity of the Lagrangian reduces to the equality of derivatives of the objective function w.r.t. p_k and p_l:

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p_k} = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}p_l} = 0 \ \Leftrightarrow \ \frac{\mathrm{d}f}{\mathrm{d}p_k} = \frac{\mathrm{d}f}{\mathrm{d}p_l} = -\mu. \tag{2.51}$$

The derivative of f w.r.t. p_k is

$$\frac{1}{2}\frac{\mathrm{d}f}{\mathrm{d}p_k} = -\frac{\pi_k}{1-p_k}\left(c - p_k\pi_k\right) + \frac{1}{1-p_k}\sum_{i=1}^{n} p_i \left(c - p_i\pi_i\right)\pi_i, \tag{2.52}$$

where $\pi_i = \pi_i(\mathbf{p}) \equiv \pi(\mathbf{p})/(1-p_i)$. Similarly we can write out the derivative w.r.t. p_l. Equating the two by further multiplying both sides by $(1-p_k)(1-p_l)$ gives

$$(1-p_l)(\eta_c - (c - p_k\pi_k)\pi_k) = (1-p_k)(\eta_c - (c - p_l\pi_l)\pi_l), \tag{2.53}$$

where $\eta_c \equiv \sum_{i=1}^n p_i(c - p_i\pi_i)\pi_i$ and hence η_c can be viewed as the expectation of a discrete random variable Z with support $\{(c-p_i\pi_i)\pi_i, i \in [n]\}$ and associated PMF $\mathbb{P}(Z = (c-p_i\pi_i)\pi_i) = p_i$ for each $i \in [n]$. So the only way to satisfy the above equality (for all k, l such that $0 < p_k < p_l$) is by requiring $(c - p_i\pi_i)\pi_i$ to be all equal for i's such that $p_i \neq 0$ (because otherwise we can always choose k', l' so that η_c lies between $(c - p_{k'}\pi_{k'})\pi_{k'}$ and $(c - p_{l'}\pi_{l'})\pi_{l'}$). In particular, $(c - p_l\pi_l)\pi_l = (c - p_k\pi_k)\pi_k$, which simplifies, after some algebra, to:

$$c = \frac{\pi(\mathbf{p})}{(1-p_k)(1-p_l)}(1 - p_k p_l). \tag{2.54}$$

Because of the constraint enforced by (2.54), we claim there are *at most two* distinct values among all the non-zero components of a potential local extremizer. To see this, we prove by contradiction. Assume there exist p_j, p_k, p_l such that $0 < p_j < p_k < p_l < 1$. Then (2.54) must hold with indices $\{j, k\}$ replacing indices $\{k, l\}$. Equating the two resulting expressions for c gives $p_j = p_l$, a contradiction.

With the above claim, we only need to consider points that have at most two distinct non-zero component values. Define $\mathcal{V}(\mathbf{p}) = \{a \in (0, 1] : \exists i \in [n] : p_i = a\}$ as the set of non-zero values taken by a $\mathbf{p} \in \partial\mathcal{S}$, and $\mathcal{Z}(\mathbf{p}) = \{i \in [n] : p_i = 0\}$ as the set of indices where $\mathbf{p}$ has a zero value. The set of probability vectors taking at most two distinct non-zero component values is then denoted $\mathcal{P}^{(2)} = \{\mathbf{p} \in \partial\mathcal{S} : |\mathcal{V}(\mathbf{p})| \in \{1, 2\}\}$. We partition this set into two subsets, $\mathcal{P}^{(2)} = \mathcal{P}_a^{(2)} \cup \mathcal{P}_b^{(2)}$, which are in turn each partitioned into two subsets, $\mathcal{P}_a^{(2)} = \mathcal{P}_{a,1}^{(2)} \cup \mathcal{P}_{a,2}^{(2)}$ and $\mathcal{P}_b^{(2)} = \mathcal{P}_{b,1}^{(2)} \cup \mathcal{P}_{b,2}^{(2)}$, where

$$\begin{aligned}
\mathcal{P}_a^{(2)} &= \{\mathbf{p} \in \mathcal{P}^{(2)} : \mathcal{Z}(\mathbf{p}) = \emptyset\} & \mathcal{P}_b^{(2)} &= \{\mathbf{p} \in \mathcal{P}^{(2)} : \mathcal{Z}(\mathbf{p}) \neq \emptyset\} \\
\mathcal{P}_{a,1}^{(2)} &= \{\mathbf{p} \in \mathcal{P}_a^{(2)} : |\mathcal{V}(\mathbf{p})| = 1\} & \mathcal{P}_{a,2}^{(2)} &= \{\mathbf{p} \in \mathcal{P}_a^{(2)} : |\mathcal{V}(\mathbf{p})| = 2\} \\
\mathcal{P}_{b,1}^{(2)} &= \{\mathbf{p} \in \mathcal{P}_b^{(2)} : |\mathcal{Z}(\mathbf{p})| = n - 1\} & \mathcal{P}_{b,2}^{(2)} &= \{\mathbf{p} \in \mathcal{P}_b^{(2)} : |\mathcal{Z}(\mathbf{p})| \in \{1, \ldots, n-2\}\}.
\end{aligned} \tag{2.55}$$

In words, $\mathcal{P}_a^{(2)}$ holds $\mathbf{p} \in \partial\mathcal{S}$ with no component equal to zero and at most two distinct (non-zero) values, while $\mathcal{P}_b^{(2)}$ holds those with at least one component equal to zero and at most two distinct non-zero values. Likewise, $\mathcal{P}_{a,1}^{(2)}$ holds $\mathbf{p}$ with no zero components and only one (non-zero) value,

meaning $\mathcal{P}^{(2)}_{a,1} = \{\frac{1}{n}\mathbf{1}\}$, and $\mathcal{P}^{(2)}_{a,2}$ holds $\mathbf{p}$ with all components taking one of two non-zero values, and both values held by some component. Finally, $\mathcal{P}^{(2)}_{b,1}$ holds $\mathbf{p}$ with all but one of the n entries holding value zero, meaning $\mathcal{P}^{(2)}_{b,1} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, and $\mathcal{P}^{(2)}_{b,2}$ holds $\mathbf{p}$ with between one and $n-2$ components taking value zero, and all non-zero components taking at most two distinct (non-zero) values. The next step (Step 3) in the proof focuses on $\mathcal{P}^{(2)}_{a,2}$, while Step 4 focuses on $\mathcal{P}^{(2)}_{b,2}$; the simpler cases $\mathcal{P}^{(2)}_{a,1}$ and $\mathcal{P}^{(2)}_{b,1}$ will be left until the end.

Step 3: any $\mathbf{p} \in \mathcal{P}^{(2)}_{a,2}$ cannot be a global maximizer. We define the subset $\mathcal{P}^{(2),*}_{a,2} \subseteq \mathcal{P}^{(2)}_{a,2}$ as the collection of points from $\mathcal{P}^{(2)}_{a,2}$ that also satisfies (2.54), which is a necessary condition for any such $\mathbf{p}$ to be a potential extremizer. In order to rule out the possibility that a point from $\mathcal{P}^{(2)}_{a,2}$ can be a global maximizer, based on the KKT condition analysis, we only need to show the original objective function f maximized over $\mathbf{p} \in \mathcal{P}^{(2),*}_{a,2}$ is no larger than say $f(\mathbf{e}_i) = d(\mathbf{c}, \mathbf{e}_i)^2$, equivalently we show another function $\tilde{f}$ maximized over $\mathbf{p} \in \mathcal{P}^{(2),*}_{a,2}$ is no larger than $f(\mathbf{e}_i)$ where $\tilde{f} = f$ for all $\mathbf{p} \in \mathcal{P}^{(2),*}_{a,2}$. It suffices to work with an enlarged feasible set, meaning we shall show $\tilde{f}$ maximized over $\mathbf{p} \in \mathcal{P}^{(2)}_{a,2}$ is still smaller than $f(\mathbf{e}_i)$.

As $\mathbf{p} \in \mathcal{P}^{(2)}_{a,2}$ by assumption, there is no loss in generality in denoting the two non-zero values it takes by $\mathcal{V}(\mathbf{p}) = \{p_s, p_l\}$ for $0 < p_s < p_l < 1$, where s stands for small and l for large (and do not denote indices). Assume there are k ($1 \leq k \leq n-1$) components that equal p_s and hence $(n-k)$ components equal p_l. Then $kp_s + (n-k)p_l = 1$, and it follows from these assumptions that $0 < p_s < \frac{1}{n} < p_l < 1$, where we emphasize the strictness of each of the above inequalities. Because of the assumption of exactly two distinct non-zero values for $\mathbf{p}$, (2.54) simplifies to

$$c = (1-p_s)^{k-1}(1-p_l)^{n-k-1}(1-p_s p_l). \tag{2.56}$$

Recall all the points from the set $\mathcal{P}^{(2)}_{a,2}$ satisfy $kp_s+(n-k)p_l = 1$, only points from the subset $\mathcal{P}^{(2),*}_{a,2}$ also satisfy (2.56). We now express the original objective function f from (2.45) as another function, $\tilde{f}(p_s, k)$, where $f(\mathbf{p}) = \tilde{f}(p_s, k)$ for all $\mathbf{p} \in \mathcal{P}^{(2),*}_{a,2}$, i.e., for all $\mathbf{p}$ for which both $kp_s + (n-k)p_l = 1$

and (2.56) hold:

$$\begin{aligned}
\tilde{f}(p_s,k) &= k\left(c-p_s(1-p_s)^{k-1}(1-p_l)^{n-k}\right)^2+(n-k)\left(c-p_l(1-p_s)^k(1-p_l)^{n-k-1}\right)^2\\
&= k\left((1-p_s)^{k-1}(1-p_l)^{n-k-1}(1-p_sp_l)-p_s(1-p_s)^{k-1}(1-p_l)^{n-k}\right)^2+\\
&\quad (n-k)\left((1-p_s)^{k-1}(1-p_l)^{n-k-1}(1-p_sp_l)-p_l(1-p_s)^k(1-p_l)^{n-k-1}\right)^2\\
&= \left((1-p_s)^{k-1}(1-p_l)^{n-k-1}\right)^2\left(k(1-p_s)^2+(n-k)(1-p_l)^2\right)\\
&= \left(\frac{c}{1-p_sp_l}\right)^2\left(n+kp_s^2+(n-k)p_l^2-2(kp_s+(n-k)p_l)\right)\\
&= c^2(n-k)\frac{(n-k)(n-2)+1-2kp_s+knp_s^2}{(n-k-p_s+kp_s^2)^2}. \qquad (2.57)
\end{aligned}$$

Fixing $p_s\in(0,\frac{1}{n})$ temporarily, we now show $\tilde{f}$ is monotone increasing in k for $k\in\{1,\ldots,n-1\}$. Denote $u=u(p_s,k)=(n-k)(n-2)+1-2kp_s+knp_s^2$ and $v=v(p_s,k)=n-k-p_s+kp_s^2$ so that

$$\tilde{f}(p_s,k)=c^2(n-k)\frac{u(p_s,k)}{v(p_s,k)^2}. \qquad (2.58)$$

It is straightforward to establish that $u\geq 0$, $v\geq 0$ under the given assumptions. Taking the derivative of $\tilde{f}$ w.r.t. k:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}k}\tilde{f}(p_s,k) &= \left(c^2/\left(n-p_s-k(1-p_s^2)\right)^3\right)\left(-uv+(n-k)v(np_s^2-2p_s-(n-2))-2(n-k)u(p_s^2-1)\right)\\
&= \left(c^2/\left(n-p_s-k(1-p_s^2)\right)^3\right)h(p_s,k) \qquad (2.59)
\end{aligned}$$

Therefore showing $\frac{\mathrm{d}\tilde{f}}{\mathrm{d}k}>0$ is equivalent to showing $h(p_s,k)>0$. Towards this, observe the third summand in $h(p_s,k)$ can be split evenly to be combined with the first and second summands, thus

$$h(p_s,k)=(1-np_s)\left(up_s+(n-k)(1-p_s)^2\right)>0. \qquad (2.60)$$

It follows that, for fixed $p_s\in(0,\frac{1}{n})$, $\tilde{f}(p_s,k)$ is maximized at $k=n-1$. The global maximum of $\tilde{f}(p_s,k)$ is obtained by further optimizing $\tilde{f}(p_s,n-1)$ over $p_s\in(0,\frac{1}{n})$. Setting $k=n-1$ in (2.57)

gives

$$\tilde{f}(p_s, n-1) = c^2(n-1)\frac{np_s^2 - 2p_s + 1}{\left((n-1)p_s^2 - p_s + 1\right)^2}, \tag{2.61}$$

for which

$$\left.\frac{\partial \tilde{f}(p_s,k)}{\partial p_s}\right|_{k=n-1} = -2\frac{c^2(n-1)^2}{v} p_s\left(np_s^2 - 3p_s + 1\right) < 0.$$

The inequality holds since the quadratic $np_s^2 - 3p_s + 1$ can be verified to be positive for $n \geq 2$ and $p_s \in (0, \frac{1}{n})$. Therefore, the maximum (indeed supremum) of $\tilde{f}(p_s, n-1)$ is obtained when $p_s \to 0$ (meaning in the limit $\mathbf{e}_i$ is the maximizer although $\mathbf{e}_i$ itself does not satisfy (2.54)), which according to (2.61) is $(n-1)c^2$. These monotonicities are illustrated in Fig. 2.4.

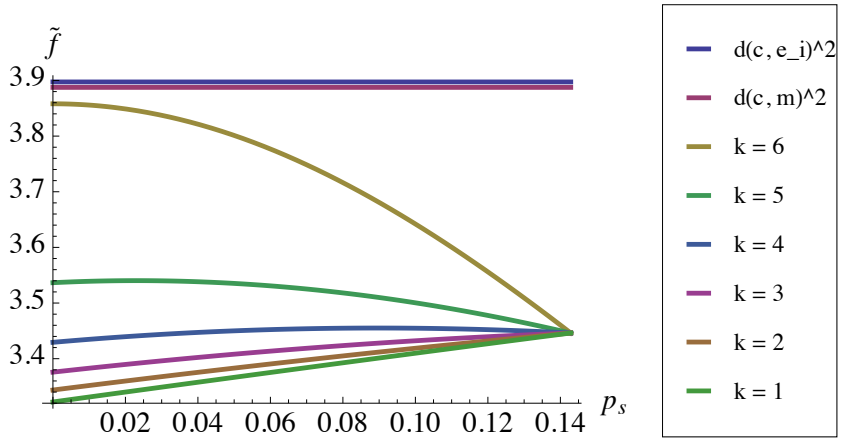

Figure 2.4: Monotonicity of function $\tilde{f}(p_s, k)$ w.r.t. k and p_s for $\mathbf{p} \in \mathcal{P}_{a,2}^{(2)}$ (Step 3) of the spherical inner bound Λ_{si} proof. Horizontal axis denotes $p_s \in (0, 1/n)$ with $n = 7$. Here $c = 0.99c_{\text{in}}^* < c_{\text{in}}^*$ so $d(\mathbf{c}, \mathbf{e}_i) > d(\mathbf{c}, \mathbf{m})$. Given p_s, $\tilde{f}$ is increasing in k; when $k = n-1$, $\tilde{f}$ is decreasing in p_s so the supreme of $\tilde{f}$ over the set $\mathcal{P}_{a,2}^{(2)}$ is achieved as $p_s \to 0$, which is $(n-1)c^2 \approx 3.85807$.

Observe when we maximize $\tilde{f}(p_s, k)$ we effectively enlarge the feasible set from $\mathcal{P}_{a,2}^{(2),*}$ to be $\mathcal{P}_{a,2}^{(2)}$ because we do not check whether (2.54) is satisfied. Recall f is identically equal to $\tilde{f}$ only for $\mathbf{p} \in \mathcal{P}_{a,2}^{(2),*}$ because $\tilde{f}$ is derived from f by applying (2.54). Therefore we have

$$f(\mathbf{p}') = \tilde{f}(\mathbf{p}') \leq \max_{\mathbf{p} \in \mathcal{P}_{a,2}^{(2),*}} \tilde{f}(\mathbf{p}) \leq \max_{\mathbf{p} \in \mathcal{P}_{a,2}^{(2)}} \tilde{f}(\mathbf{p}) = (n-1)c^2, \ \forall \mathbf{p}' \in \mathcal{P}_{a,2}^{(2),*}. \tag{2.62}$$

Summarizing, so far we have shown, suppose there exists a potential extremizer $\mathbf{p}$ whose com-

ponents are all non-zero but not all identical, then in order to satisfy the first-order KKT necessary conditions, the original objective function f evaluated at such a point is *upper bounded* by $\tilde{f}(0, n-1) = (n-1)c^2$. Now since $f(\mathbf{e}_i) = d(\mathbf{c}, \mathbf{e}_i)^2 = (n-1)c^2 + (c-1)^2$, this means no $\mathbf{p} \in \mathcal{P}_{a,2}^{(2)}$ can achieve a higher objective value than $\mathbf{e}_i$ does (i.e., case $\mathcal{P}_{b,1}^{(2)}$) in terms of globally maximizing the original objective function. In fact, this property does not depend on choosing the thresholding $c_{\text{in}}^* = (1 - nm^2)/(2(1 - nm))$. This property is useful in Step 4 below.

Step 4: any $\mathbf{p} \in \mathcal{P}_{b,2}^{(2)}$ cannot be a global maximizer. Fix $\mathbf{p} \in \mathcal{P}_{b,2}^{(2)}$ and let $s = n - |\mathcal{Z}(\mathbf{p})| \in \{2, \ldots, n-1\}$ be the number of non-zero components. Evaluating the original objective function, (2.45), for such a point yields

$$f(\mathbf{p}) = (n-s)(c-0)^2 + \sum_{i=1}^{s} \left(c - p_i \prod_{j \neq i}^{s} (1 - p_j) \right)^2 . \tag{2.63}$$

For each given s, $f(\mathbf{p})$ is maximized iff the second summand above is maximized. Maximizing the above second summand can be thought of as performing the same optimization problem in an s-dimensional space where the s-vector $\mathbf{p}$ duplicates all the s non-zero components from the original n-vector $\mathbf{p}$. Then, one may view this s-vector with no zero component as a member of $\mathcal{P}_a^{(2)}$, but with the dimension reduced from n to s. There are two possibilities: this point is either in $\mathcal{P}_{a,2}^{(2)}$ or in $\mathcal{P}_{a,1}^{(2)}$.

Consider the first possibility, i.e., $\mathcal{P}_{a,2}^{(2)}$. Based on the analysis of this case in Step 3 (with the dimension reduced from n to s), and the upper bound (2.62) in particular, it follows that

$$\sum_{i=1}^{s} \left(c - p_i \prod_{j \neq i}^{s} (1 - p_j) \right)^2 \leq (s-1)c^2,$$

and hence $f(\mathbf{p}) \leq (n-s)(c-0)^2 + (s-1)c^2 = (n-1)c^2$, which is the same upper bound for candidates in case $\mathcal{P}_{a,2}^{(2)}$ in the original n-dimensional space. It follows that, in this reduced dimensional space, points in $\mathcal{P}_{a,2}^{(2)}$ cannot achieve a higher objective value than that achieved by the points $\mathbf{e}_i$ in the original space.

Consider the second possibility, i.e., $\mathcal{P}_{a,1}^{(2)}$, namely the all-rates-equal point in this s-dimensional

space. There are two subcases: *i)* $c \leq c^*_{\text{in}}(s) = (1 - sm^2(s))/(2(1 - sm(s)))$, and *ii)* $c > c^*_{\text{in}}(s)$. Note we write $c^*_{\text{in}}(s)$ to highlight it is a function of s, the corresponding dimension. Case *i)* can be skipped, due to (2.42) and the observation that the $\mathbf{e}_i$ in this s-dimensional space is also the $\mathbf{e}_i$ in the original n-dimensional space. Recall, $\mathbf{e}_i$ (in the set $\mathcal{P}^{(2)}_{b,1}$) will be addressed in the final step. For case *ii)* we now directly show the all-rates-equal point in this s-dimensional space cannot achieve a higher objective value than that in the original space. First, it is straightforward to establish the inequality

$$n\left(c - \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1}\right)^2 = f(\mathbf{m}(n)) \geq f(\mathbf{m}(s)) = (n-s)c^2 + s\left(c - \frac{1}{s}\left(1 - \frac{1}{s}\right)^{s-1}\right)^2 \tag{2.64}$$

holds iff

$$2c(sm(s) - nm(n)) \geq sm^2(s) - nm^2(n). \tag{2.65}$$

Since $c > c^*_{\text{in}}(s)$, (2.65) is equivalent to

$$sm(s)(1 - m(s)) - nm(n)(1 - m(n)) + snm(s)m(n)(m(s) - m(n)) \geq 0. \tag{2.66}$$

Since $snm(s)m(n)(m(s) - m(n)) \geq 0$, to show (2.66) it suffices to show the function $g(n) \equiv nm(n)(1 - m(n))$ is monotone decreasing in n for $n \geq 3$ (recall $2 \leq s \leq n-1$). Towards this we find the derivative of $g(n)$ as

$$\begin{aligned}\frac{\mathrm{d}g(n)}{\mathrm{d}n} &= \frac{1}{2n^2}\left(1 - \frac{1}{n}\right)^{n-1}\left[-2\left(1 - \frac{1}{n}\right)^{n-1} + 2n + 2n^2\left(1 - \frac{2}{n}\left(1 - \frac{1}{n}\right)^{n-1}\right)\log\left(1 - \frac{1}{n}\right)\right] \\ &= \frac{1}{2n^2}\left(1 - \frac{1}{n}\right)^{n-1}\tilde{g}(n).\end{aligned} \tag{2.67}$$

It now suffices to show $\tilde{g}(n) < 0$. For $n = 3, 4$ this can be verified; for $n \geq 5$ we apply the inequality $\log(1 + x) \leq x - \frac{x^2}{2}$ for all $x \in (-1, 0]$ and get

$$\tilde{g}(n) \leq 2\left(1 - \frac{1}{n}\right)^{n-1}\left(1 + \frac{1}{n}\right) - 1 \leq 0, \tag{2.68}$$

where the inequality follows from the monotonicity in n of the upper bound on $\tilde{g}(n)$. Therefore we have shown the desired inequality (2.66). This means that, even if a point is from $\mathcal{P}_{a,1}^{(2)}$, it cannot be a global maximizer as it cannot achieve a higher objective value than $\mathbf{e}_i$ (case i)) and/or $\mathbf{m}$ (case ii)) in the original n-dimensional space. This concludes Step 4.

Finally, we are left with only cases $\mathcal{P}_{a,1}^{(2)}$ and $\mathcal{P}_{b,1}^{(2)}$. We can verify by checking the KKT conditions that $\mathbf{m}$ is always eligible to be a local extremizer, while $\mathbf{e}_i$ is eligible to be a local maximizer iff $c \leq 1$. Therefore, we conclude the global maximum of the original optimization problem can be obtained by evaluating and comparing f at two points $\mathbf{m}$, $\mathbf{e}_i$. Furthermore, recall the objective function is defined as $d(\mathbf{c}, \mathbf{x})^2$ for $\mathbf{x} \in \partial\Lambda$. Then as a consequence of (2.42), we can actually conclude in a more general manner: the global maximum occurs at i) any $\mathbf{e}_i$ when $c < c_{\text{in}}^*$, ii) any $\mathbf{e}_i$ and $\mathbf{m}$ when $c = c_{\text{in}}^*$, and iii) $\mathbf{m}$ when $c > c_{\text{in}}^*$. See Fig. 2.5 for an illustration. For Λ_{si}^*, the global maximizers are both $\mathbf{e}_i$ and $\mathbf{m}$ giving the maximum of f as $(n-1)\, c_{\text{in}}^{*2} + (c_{\text{in}}^* - 1)^2 = n\,(c_{\text{in}}^* - m)^2$, as desired in (2.41). □

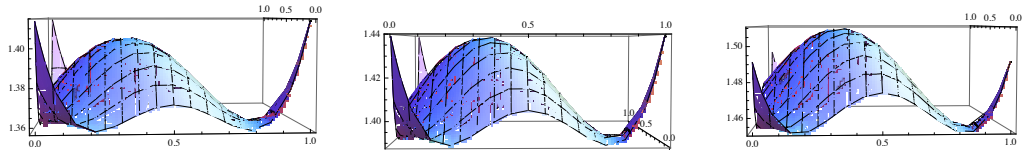

Figure 2.5: Original objective function $f(\mathbf{p})$ for Λ_{si} when $n = 3$ for various c. Horizontal axes are p_2, p_3. **Left:** $c = 0.99 c_{\text{in}}^*$ with global maximizers: $\mathbf{e}_i$. **Middle:** $c = c_{\text{in}}^*$, with global maximizers: $\mathbf{e}_i$ & $\mathbf{m}$. **Right:** $c = 1.02 c_{\text{in}}^*$, with global maximizer:$\mathbf{m}$.

Remark 3. *Since the hyperplane inducing the optimal polyhedral inner bound Λ_{pi}^* is a supporting hyperplane of the convex body $\mathcal{B}(\mathbf{c}, r_{\text{in}}(c))$, due to the constructions of Λ_{pi}^* and Λ_{si} it follows that Λ_{si} is always tighter than Λ_{pi}^*.*

We now proceed to the spherical outer bound. There are many similarities between the definitions, propositions and proof techniques for the spherical inner and outer bounds. In both cases there exists a set inclusion relationship which implies the optimal bound arises when c is chosen to be the minimum possible.

Definition 8. $\Lambda_{\rm so}(c) \equiv \mathcal{S} \setminus \mathcal{B}(\mathbf{c}, r_{\rm out}(c))$*, where the center of the ball* $\mathbf{c} = c\mathbf{1}$ *for all* $c \geq 1$*, and its radius* $r_{\rm out}(c) \equiv d(\mathbf{c}, \mathbf{e}_i) = \sqrt{(c-1)^2 + (n-1)c^2}$.

Proposition 10 (spherical outer bound)**.** *For each* $c \geq 1$*, the set* $\Lambda_{\rm so}(c)$ *is an outer bound on* Λ *for* $n \geq 2$*. Among these, the tightest is given when* $c = c^*_{\rm out} = 1$*:*

$$\Lambda^*_{\rm so} = \Lambda_{\rm so}(c^*_{\rm out}) = \left\{\mathbf{x} \in \mathcal{S} : \|\mathbf{x} - \mathbf{1}\| \geq \sqrt{n-1}\right\}. \tag{2.69}$$

Proof. By Def. 8, in order to induce an outer bound we must have $c > m$. Moreover, $d(\mathbf{c}, \mathbf{m}) > d(\mathbf{c}, \mathbf{e}_i) = r_{\rm out}$, which is equivalent to $c > (1 - nm^2)/(2(1 - nm))$. This means there remain two possible intervals for c: *i*) $(1-nm^2)/(2(1-nm)) < c < 1$ and *ii*) $c \geq 1$. For each one we investigate whether a ball with parameter c in that interval induces a valid outer bound on Λ.

For case *i*), we can compute that $\mathbf{e}_1, \mathbf{x}_b \in \partial\mathcal{B}(\mathbf{c}, r_{\rm out})$ for $\mathbf{x}_b \equiv (2c-1)\mathbf{e}_1$, i.e., the boundary of the ball intersects the first coordinate axis at these two points. As $c > (1 - nm^2)/(2(1 - nm)) > \frac{1}{2}$, the line segment $\overline{\mathbf{x}_b\mathbf{e}_1} \subseteq \Lambda$ due to Λ's coordinate convexity. Furthermore, since the open line segment $\overline{\mathbf{x}_b\mathbf{e}_1} \subseteq \mathcal{B}(\mathbf{c}, r_{\rm out})$, we find $\Lambda \not\subseteq \mathcal{S} \setminus \mathcal{B}(\mathbf{c}, r_{\rm out})$ namely $\mathcal{B}(\mathbf{c}, r_{\rm out})$ does not induce a valid outer bound.

It remains to investigate case *ii*). In the rest of this proof we first show every c in this category gives a valid outer bound and, furthermore, $c = c^*_{\rm out} = 1$ yields the tightest bound. We first show $c = c^*_{\rm out} = 1$ yields the tightest bound, i.e., we show $\Lambda^*_{\rm so} \subseteq \Lambda_{\rm so}(c)$ for each $c \geq 1$. Note the equivalence

$$\begin{aligned} \Lambda^*_{\rm so} \subseteq \Lambda_{\rm so}(c) \quad &\iff \quad \mathcal{S} \setminus \left(\mathcal{B}(\mathbf{1}, \sqrt{n-1}) \cap \mathcal{S}\right) \subseteq \mathcal{S} \setminus \left(\mathcal{B}(\mathbf{c}, r_{\rm out}(c)) \cap \mathcal{S}\right) \\ &\iff \quad \mathcal{B}(\mathbf{c}, r_{\rm out}(c)) \cap \mathcal{S} \subseteq \mathcal{B}(\mathbf{1}, \sqrt{n-1}) \cap \mathcal{S}. \end{aligned} \tag{2.70}$$

Therefore we seek to prove: $\forall \mathbf{x} \in \mathcal{S}$, if $d(\mathbf{c}, \mathbf{x})^2 < r_{\rm out}(c)^2$ (i.e., $\mathbf{x} \in \mathcal{B}(\mathbf{c}, r_{\rm out}(c)) \cap \mathcal{S}$), then $d(\mathbf{1}, \mathbf{x})^2 < \left(\sqrt{n-1}\right)^2$ (i.e., $\mathbf{x} \in \mathcal{B}(\mathbf{1}, \sqrt{n-1}) \cap \mathcal{S}$). Suppose $\mathbf{x} \in \mathcal{S}$ is such that $d(\mathbf{c}, \mathbf{x})^2 < r_{\rm out}(c)^2$. Then, we

compute:

$$\begin{aligned} d(\mathbf{1},\mathbf{x})^2 &= \sum_{i=1}^n (1-x_i)^2 = \sum_{i=1}^n \left((c-x_i)-(c-1)\right)^2 = \sum_{i=1}^n (c-x_i)^2 + n(c-1)^2 - 2(c-1)\left(nc - \sum_{i=1}^n x_i\right) \\ &< (c-1)^2 + (n-1)c^2 + n(c-1)^2 - 2(c-1)(nc-1) = n-1, \end{aligned} \tag{2.71}$$

where the inequality follows from $d(\mathbf{c},\mathbf{x})^2 < r_{\text{out}}(c)^2$ and $\sum_{i=1}^n x_i \leq 1$. This shows the desired set inclusion, meaning Λ_{so}^* is the smallest set among $\{\Lambda_{\text{so}}(c), c \geq 1\}$.

It remains to show that Λ_{so}^* is a valid outer bound on Λ. For any $\mathbf{x} \in \Lambda = \Lambda_{\text{eq}}$ we must show $\mathbf{x} \in \Lambda_{\text{so}}^*$, namely $\mathbf{x}$ is outside the open ball $\mathcal{B}(\mathbf{c},1)$, or equivalently $\min_{\mathbf{x}\in\Lambda} \|\mathbf{1}-\mathbf{x}\|^2 \geq n-1$. This latter expression may be cast as an optimization problem w.r.t. $\mathbf{p}$:

$$\min_{\mathbf{p}\in[0,1]^n} f(\mathbf{x}(\mathbf{p})) \equiv \sum_{i=1}^n \left(1 - p_i \prod_{j\neq i}(1-p_j)\right)^2 \geq n-1. \tag{2.72}$$

Observe if any component of $\mathbf{p}$ equals 1 or if $\mathbf{p} = \mathbf{0}$ then $f \geq n-1$ immediately holds. So below we assume $\mathbf{p} < \mathbf{1}$ and $\mathbf{p}$ has non-zero component(s). Recall we defined $\mathcal{V}(\mathbf{p}) = \{a \in (0,1] : \exists i \in [n] : p_i = a\}$ in the proof of Prop. 9 as the set of non-zero values taken by a vector $\mathbf{p}$. We now categorize based on how many distinct non-zero values the components of $\mathbf{p}$ assume: *a)* $|\mathcal{V}(\mathbf{p})| > 1$ or *b)* $|\mathcal{V}(\mathbf{p})| = 1$.

Consider first case *a)* ($|\mathcal{V}(\mathbf{p})| > 1$), i.e., $\mathbf{p}$ has two or more distinct non-zero component values, say $0 < p_k < p_l < 1$. We will show that all such $\mathbf{p}$'s cannot be local extremizers due to the violation of Karush-Kuhn-Tucker (KKT) conditions required for optimality (note regularity is guaranteed in this case). An equivalent form of the KKT stationarity condition is that

$$\frac{1}{2}\left(\frac{\partial f}{\partial p_k} - \frac{\partial f}{\partial p_l}\right)(1-p_k)(1-p_l) = 0,\ 0 < p_k < p_l < 1, \tag{2.73}$$

which, after some algebra, may be shown to be equivalent to:

$$(1-p_l)\left(\eta - \pi_k(1-p_k\pi_k)\right) = (1-p_k)\left(\eta - \pi_l(1-p_l\pi_l)\right), \tag{2.74}$$

where $\eta \equiv \sum_i \pi_i(1-p_i\pi_i)p_i$, $\pi_i = \pi_i(\mathbf{p}) \equiv \pi(\mathbf{p})/(1-p_i)$. Note η can be interpreted as the expectation of a discrete random variable Z with support $\{\pi_i(1-p_i\pi_i), i \in [n]\}$ and associated PMF $\mathbb{P}(Z = \pi_i(1-p_i\pi_i)) = p_i$ for each $i \in [n]$. Therefore, stationarity will not be satisfied as long as we can choose indices k', l' such that $0 < p_{k'} < p_{l'} < 1$, and η lies strictly between $\pi_{k'}(1-p_{k'}\pi_{k'})$ and $\pi_{l'}(1-p_{l'}\pi_{l'})$. But, we *can* always find such indices since, following Lem. 3 (which can be easily verified), we can show the ordering: $\pi_{k'}(1-p_{k'}\pi_{k'}) < \pi_{l'}(1-p_{l'}\pi_{l'})$. This rules out the possibility that an extremizer can come from case *a*).

Consider next case *b*) ($|\mathcal{V}(\mathbf{p})| = 1$), i.e., $\mathbf{p}$ has only one distinct non-zero component value (we call such $\mathbf{p}$ "quasi-uniform"). Lem. 4 below states that for any such $\mathbf{p}$, the objective $f(\mathbf{p}) \geq n-1$, the desired lower bound in (2.72).

These two cases establish the validity of the inequality (2.72), and thereby establish the fact that Λ^*_{so} is a valid outer bound for Λ. □

The following two lemmas are used in the preceding proof of Prop. 10.

Lemma 3. *If two non-zero components of* $\mathbf{p}$ *satisfy* $p_l > p_k$*, then for* $\mathbf{x} = \mathbf{x}(\mathbf{p})$ *as defined in* (2.9)

$$x_l > x_k, \quad \pi_l > \pi_k, \quad \frac{x_l}{x_k} > \frac{p_l}{p_k}, \quad x_l - x_k < p_l - p_k. \tag{2.75}$$

Proof. Omitted. □

Lemma 4. *Fix* $t \in (0,1]$ *and* $k \in [n]$*. Suppose* $\mathbf{p} < \mathbf{1}$ *(*$\mathbf{p} \neq \mathbf{0}$*) takes only one non-zero value (i.e.,* $|\mathcal{V}(\mathbf{p})| = 1$*, and* $p_i \in \{0, t\}$ *for* $i \in [n]$*), and this value is taken by* k *components of* $\mathbf{p}$*. Then* $f(\mathbf{p}) \geq n-1$ *(for* f *in* (2.72)*), with equality iff* $k = t = 1$*, i.e.,* $f(\mathbf{p}) = n-1$ *iff* $\mathbf{p} \in \{\mathbf{e}_i\}_{i=1}^n$*.*

Proof. Wlog let $\mathbf{p} = t \cdot \sum_{i=1}^k \mathbf{e}_i$ for $k \in [n]$, $t \in (0,1]$. Substitution of such a $\mathbf{p}$ into (2.72) yields the following inequality

$$t(1-t)^{k-1} \leq 1 - \sqrt{1-1/k}, \tag{2.76}$$

meaning the lemma will be established if we can show (2.76) holds for all valid (t,k), and holds with equality iff $k = t = 1$. The inequality (2.76) is easily verified to hold strictly for *a*) $k = 1 \neq t$ and

b) $t = 1 \neq k$. If $k = t = 1$ (namely $\mathbf{p} = \mathbf{e}_1$) the original objective function in (2.72) evaluates to $n - 1$, the desired minimum. It remains to study the case $k \in \{2, \ldots, n\}$ and $0 < t < 1$. Define $g(t) \equiv t(1-t)^{k-1}$. The only stationary point of g on $t \in [0, 1)$ is $t^* = 1/k$, at which the second derivative can be verified to be strictly negative, meaning t^* is the unique maximizer. And hence we need to show (2.76) when $t = 1/k$, namely $(1 - 1/k)^{k-1} \leq k\left(1 - \sqrt{1 - 1/k}\right)$. The derivative of its LHS can be shown to be negative using the inequality $\log(1 + x) \leq x$ for $x > -1$. Thus the sequence $\{(1 - 1/k)^{k-1}\}$ is upper bounded by $(1 - 1/2)^{2-1} = 1/2$. On the other hand, using AM-GM inequality $\sqrt{1 - 1/k} < (1 - 1/k + 1)/2$, one can see the RHS is strictly lower bounded by $1/2$. This shows the desired inequality (2.76), thus proving this lemma. □

Remark 4. *The Cauchy-Schwarz inequality gets close to proving the desired inequality* (2.72)*, but is insufficient by itself (note* $\sum_i x_i \leq 1$ *as* $\Lambda \subseteq \mathcal{S}$*):*

$$\sum_{i=1}^{n} (1 - x_i)^2 \geq \frac{\left(\sum_{i=1}^{n} 1 \cdot (1 - x_i)\right)^2}{\sum_{i=1}^{n} 1^2} = \frac{\left(n - \sum_{i=1}^{n} x_i\right)^2}{n} \geq \frac{(n-1)^2}{n}, \tag{2.77}$$

which is slightly weaker than the bound of $n - 1$ *required to show* (2.72)*.*

The optimal spherical inner and outer bounds Λ^*_{si}, Λ^*_{so} together with $\partial\Lambda$ are shown in Fig. 2.6 for $n = 2$ and 3.

It seems hard to obtain the volume of these spherical bounds in closed-form for arbitrary n. Essentially, the problem is one of integrating over the intersection between a (solid) hypersphere and $[0, 1]^n$. It is natural to attempt to bound the volume. Below, we illustrate such an attempt using Λ^*_{so} as an example.

We take a probabilistic approach. As the volume of the unit box $[0, 1]^n$ always equals 1, and as $\Lambda^*_{\text{so}} \subseteq [0, 1]^n$, it follows that its volume can be interpreted as the probability that a point uniformly distributed over $[0, 1]^n$ falls into the set Λ^*_{so}. More precisely, for i.i.d. Unif$[0, 1]$ random variables

(RV) $y_1, \ldots, y_n$,

$$\begin{aligned} \mathrm{vol}(\Lambda^*_{\mathrm{so}}) &= \mathbb{P}\left(\mathbf{y} \in \mathcal{S},\ \mathbf{y} \notin \mathcal{B}(\mathbf{1}, \sqrt{n-1})\right) \overset{(a)}{=} \mathbb{P}\left(\mathbf{y} \notin \mathcal{B}(\mathbf{1}, \sqrt{n-1})\right) \\ &= \mathbb{P}\left(\sum_i (1-y_i)^2 \geq n-1\right) \overset{(b)}{=} \mathbb{P}\left(\sum_i y_i^2 \geq n-1\right), \end{aligned} \tag{2.78}$$

where (a) follows from Lem. 5 given below and (b) is due to the observation that $1 - y_i$ are i.i.d. Unif$[0, 1]$ RV's too. We note the *uniform sum distribution* (also known as *Irwin-Hall distribution*), $\sum_i y_i$, has a known closed-form density function, yet this does not seem to be the case for $\sum_i y_i^2$. Then one natural thing to do is to bound this tail probability. A typical form of the Chernoff bound states that for a random variable Z (usually expressed as a sum of independent RVs), an upper bound on the (upper) tail probability is $\mathbb{P}(Z \geq t) \leq \inf_{s \geq 0} \mathrm{e}^{-st} \mathbb{E}\left[\mathrm{e}^{sZ}\right]$. Substituting $\sum_i y_i^2$ for Z yields:

$$\mathrm{vol}(\Lambda^*_{\mathrm{so}}) \leq \inf_{s \geq 0} \mathrm{e}^{-s(n-1)} \mathbb{E}\left[\mathrm{e}^{s \sum_i y_i^2}\right] = \inf_{s \geq 0} \mathrm{e}^{-s(n-1)} \prod_i \mathbb{E}\left[\mathrm{e}^{s y_i^2}\right] = \inf_{s \geq 0} \mathrm{e}^{-s(n-1)} \left(\int_0^1 \mathrm{e}^{s y_i^2} \mathrm{d}y_i\right)^n. \tag{2.79}$$

The minimizer s^* is hard to be obtained in closed-form. Worse still, the numerically optimized upper bound is not close to the actual tail probability (i.e., the volume of Λ^*_{so}).

Lemma 5. $[0, 1]^n \setminus \mathcal{S} \subseteq \mathcal{B}\left(\mathbf{1}, \sqrt{n-1}\right)$. *In words this says the unit box with the unit simplex subtracted lies completely inside the ball* $\mathcal{B}\left(\mathbf{1}, \sqrt{n-1}\right)$.

Proof. Given $\mathbf{x} \in [0, 1]^n, \sum_i x_i > 1$, we need to show $\sum_i (1 - x_i)^2 < n - 1$, which is easily verifiable since

$$\sum_i (1-x_i)^2 = n - 2\sum_i x_i + \sum_i x_i^2 \leq n - 2\sum_i x_i + \sum_i x_i = n - \sum_i x_i < n - 1. \tag{2.80}$$

Note this lemma can be equivalently stated as $[0, 1]^n \setminus \mathcal{B}\left(\mathbf{1}, \sqrt{n-1}\right) \subseteq \mathcal{S}$ which implies if a point is from $[0, 1]^n$ but not in $\mathcal{B}\left(\mathbf{1}, \sqrt{n-1}\right)$ then it's guaranteed to be in $\mathcal{S}$. This observation is used step (a) in (2.78). □

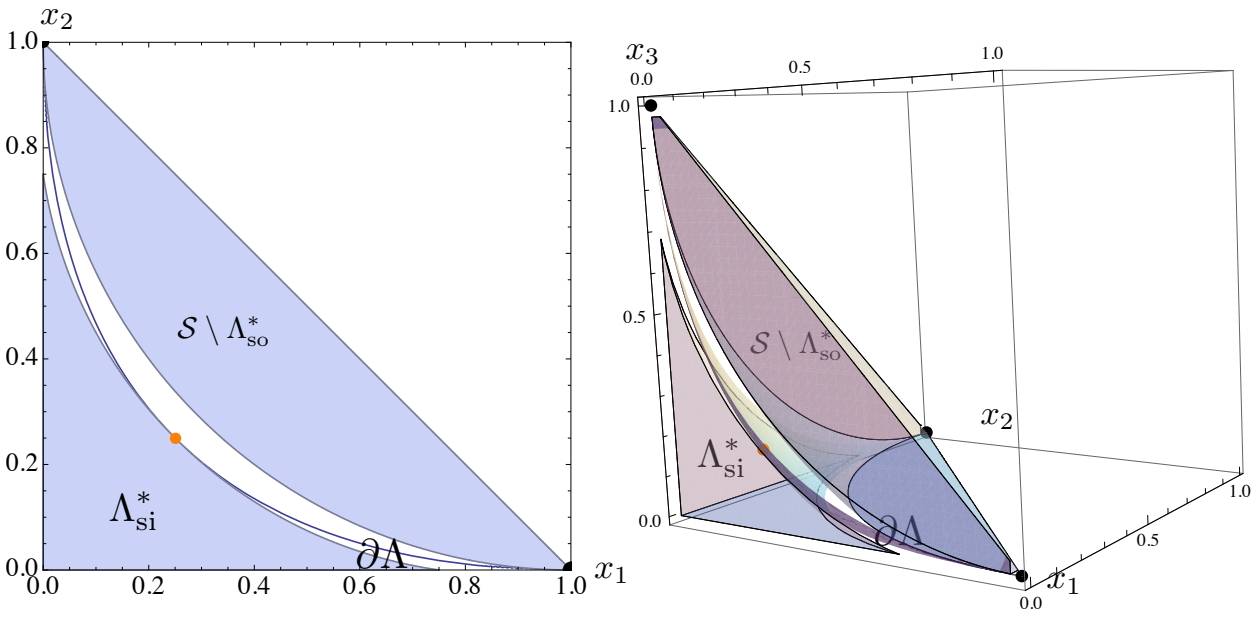

Figure 2.6: Optimal spherical bounds for $n = 2$ (left) and 3 (right): inner bound and the complement (w.r.t. $\mathcal{S}$) of outer bound are shown in solid; in between sits $\partial\Lambda$. Also shown are the all-rates-equal point $\mathbf{m}$ (orange) and corner points $\mathbf{e}_i$'s (black).

2.6 Ellipsoid inner and outer bounds on Λ

We now turn to the third, and final, class of bounds on Λ. In this section we establish inner and outer bounds, each induced by a parameterized family of ellipsoids. This section is organized into three subsections. First, in §2.6.1 we prove three results: *i*) the set of ellipsoids that inherit all the permutation symmetries of Λ are characterized by three scalars (c, a_1, a_2) (Prop. 11), *ii*) the sufficiency of working only with $\partial\Lambda$ for the purpose of proving the correctness of the induced bound (Props. 14 and 15), and *iii*) a property of a local extremizer from $\partial\Lambda$ (Prop. 16). Next, in §2.6.2 we present the parameterized families of ellipsoid inner (Prop. 18) and outer (Prop. 17) bounds. The derivation is based on the Karush-Kuhn-Tucker (KKT) optimality conditions. Finally, in §2.6.3, we provide an alternative proof of the ellipsoid outer bound by working in a transformed space and leveraging Schur-convexity.

2.6.1 Simplification of parameter space

The contributions of this section were outlined above. We consider open ellipsoids $\mathcal{E}$ of the form ([23]):

$$\mathcal{E} = \left\{\mathbf{x} : (\mathbf{x}-\mathbf{c})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{x}-\mathbf{c}) < 1\right\}. \tag{2.81}$$

Here $\mathbf{c}$ is the center of the ellipsoid and the $n \times n$ symmetric and positive definite matrix $\mathbf{R}$ has the spectral decomposition $\mathbf{R} = \mathbf{Q}\mathbf{D}\mathbf{Q}^{\mathsf{T}}$ where $\mathbf{Q} = [\mathbf{q}_1 \cdots \mathbf{q}_n]$ is orthonormal and holds the eigenvectors of $\mathbf{R}$ (which are the directions of the n axes of the ellipsoid), and $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ holds the eigenvalues of $\mathbf{R}$. Each $a_i = \sqrt{\lambda_i}$ is the semi-axis length in the direction $\mathbf{q}_i$. Denote the boundary of $\mathcal{E}$ by $\partial\mathcal{E} = \left\{\mathbf{x} : (\mathbf{x}-\mathbf{c})^{\mathsf{T}}\mathbf{R}^{-1}(\mathbf{x}-\mathbf{c}) = 1\right\}$.

Our approach is to approximate the surface $\partial\Lambda$ with part of the surface of an ellipsoid, and then form inner and outer bounds on Λ by subtracting these ellipsoids from the unit simplex $\mathcal{S}$. This is the same approach that was used in constructing the spherical bounds (§2.5). More concretely, we want to find inner and outer bounding ellipsoids $\mathcal{E}_{\mathrm{in}}, \mathcal{E}_{\mathrm{out}}$ such that $\Lambda_{\mathrm{ei}} \subseteq \Lambda \subseteq \Lambda_{\mathrm{eo}}$, where $\Lambda_{\mathrm{ei}} \equiv \mathcal{S} \setminus \mathcal{E}_{\mathrm{in}},\ \Lambda_{\mathrm{eo}} \equiv \mathcal{S} \setminus \mathcal{E}_{\mathrm{out}}$.

Although we are not able to characterize them further, we define the optimal inner and outer bounding ellipsoids:

$$\mathcal{E}_{\mathrm{in}}^{*} \equiv \underset{\mathcal{E}_{\mathrm{in}}:\overline{\mathcal{E}}_{\mathrm{in}}\cap\mathcal{S}\supseteq\overline{\Lambda^c}\cap\mathcal{S}}{\arg\min}\ \mathrm{vol}(\mathcal{E}_{\mathrm{in}} \cap \mathcal{S}) = \underset{\mathcal{E}_{\mathrm{in}}:\overline{\mathcal{E}}_{\mathrm{in}}\cap\mathcal{S}\supseteq\partial\Lambda}{\arg\min}\ \mathrm{vol}(\mathcal{E}_{\mathrm{in}} \cap \mathcal{S}) \tag{2.82}$$

$$\mathcal{E}_{\mathrm{out}}^{*} \equiv \underset{\mathcal{E}_{\mathrm{out}}:\mathcal{E}_{\mathrm{out}}\cap\Lambda=\emptyset}{\arg\min}\ \mathrm{vol}(\mathcal{S} \setminus \mathcal{E}_{\mathrm{out}}) = \underset{\mathcal{E}_{\mathrm{out}}:\mathcal{E}_{\mathrm{out}}\cap\Lambda=\emptyset}{\arg\max}\ \mathrm{vol}(\mathcal{E}_{\mathrm{out}} \cap \mathcal{S}), \tag{2.83}$$

where the second equality in (2.82) follows from Lem. 11.

A result in convex geometry states that for any convex body there exists a unique maximum (resp. minimum) volume inscribed (resp. circumscribing) ellipsoid, called the *Löwner-John ellipsoid*. Although we have a convex body $\Lambda^c \cap \mathcal{S}$, our objective is *not* to identify an inscribed/circumscribing ellipsoid with extremized volume for this set. Rather, our figure of merit (in (2.82) and (2.83)) is to extremize the volume of the *intersection* between the ellipsoid and the simplex. For example, our $\mathcal{E}_{\mathrm{out}}$ need not lie *entirely* within the convex body.

In general, analytical characterization of the Löwner-John ellipsoid is hard (see e.g.,[24 23]). One constructive result, though, is that the Löwner-John ellipsoid is an *invariant ellipsoid*, meaning it inherits *all* the symmetries of the convex body[24]. The intuition is that if there were some symmetry that the volume optimal ellipsoid is not endowed with, then using that particular symmetry one can construct another distinct volume optimal ellipsoid, hence contradicting the uniqueness of the Löwner-John ellipsoid.

In the spirit of the above result, we restrict our attention to ellipsoids that inherit all the symmetries of the convex body $\overline{\Lambda^c} \cap \mathcal{S}$. Note $\overline{\Lambda^c} \cap \mathcal{S}$, Λ and $\mathcal{S}$ all have full permutation symmetry. Prop. 11 below states some consequences of inheriting this permutation symmetry. Its proof uses the following three lemmas. The proof of Lem. 6 is given in §2.8.2 and the proof of Lem. 8 is omitted as it is easy to verify.

Lemma 6. *Fix an ellipsoid* $\mathcal{E} = \left\{\mathbf{x} : (\mathbf{x}-\mathbf{c})^\mathsf{T}\mathbf{R}^{-1}(\mathbf{x}-\mathbf{c}) < 1\right\}$ *in* $\mathbb{R}^n$ *with center* $\mathbf{c}$*. A hyperplane passing through* $\mathbf{c}$ *with its normal vector being one of the axes/eigenvectors of* $\mathbf{R}$ *is a reflecting hyperplane for* $\mathcal{E}$*, i.e.,* $\mathcal{E}$ *is symmetric w.r.t. this hyperplane. Conversely, the normal vector of any reflecting hyperplane of* $\mathcal{E}$ *can be considered as an axis/eigenvector of* $\mathbf{R}$*.*

Note that in the context of a full-dimensional ellipsoid, the axis (direction), eigenvector, and reflecting hyperplane's normal vector are all essentially the same thing.

Lemma 7. *For a symmetric matrix* $\mathbf{R}$*, if the two eigenvalues associated with two of the eigenvectors of* $\mathbf{R}$ *are distinct, then these two eigenvectors must necessarily be orthogonal (not just linearly independent).*

Proof. Suppose eigenvectors $\mathbf{v}, \mathbf{w}$ have associated distinct eigenvalues λ_v, λ_w, meaning $\mathbf{R}\mathbf{w} = \lambda_w\mathbf{w}$, $\mathbf{R}\mathbf{v} = \lambda_v\mathbf{v}$. First we write $\mathbf{w}^\mathsf{T}\mathbf{R}\mathbf{v} = \lambda_v\mathbf{w}^\mathsf{T}\mathbf{v}$. Next since $\mathbf{R}$ is symmetric, we can also write $\mathbf{w}^\mathsf{T}\mathbf{R}\mathbf{v} = (\mathbf{R}\mathbf{w})^\mathsf{T}\mathbf{v} = \lambda_w\mathbf{w}^\mathsf{T}\mathbf{v}$. These two give $(\lambda_w - \lambda_v)\mathbf{w}^\mathsf{T}\mathbf{v} \equiv 0$. As $\lambda_w \neq \lambda_v$, it has to hold that $\mathbf{w}^\mathsf{T}\mathbf{v} = 0$. □

Lemma 8. *The linear combination of some eigenvectors associated with the same eigenvalue is also an eigenvector (with the same eigenvalue). In fact, all such eigenvectors are in the same eigen-*

subspace.

Proposition 11. *The class of ellipsoids invariant under permutations of the coordinate axes is the set of ellipsoids parameterized by* $(c, a_1, a_2) \in \mathbb{R}^3_+$ *with the properties that*

1. *the center is at* $\mathbf{c} = c\mathbf{1}$
2. *one axis is along the all-rates-equal ray with direction* $\mathbf{1}$ *and has semi-axis length* a_1
3. *the* $n-1$ *remaining axes are arbitrary (provided they, together with the axis aligned with* $\mathbf{1}$*, form an orthonormal set) and have common semi-axis lengths* a_2*.*

Proof. A set $S \subseteq \mathbb{R}^n$ is invariant under permutations if for any permutation σ of $[n]$ we have $\mathbf{x} = (x_1, \ldots, x_n) \in S$ iff $\sigma(\mathbf{x}) = (x_{\sigma(1)}, \ldots, x_{\sigma(n)}) \in S$. It is straightforward to see that any ellipsoid parameterized by (c, a_1, a_2) obeying the properties in the proposition is in fact permutation invariant. It remains to show each invariant ellipsoid has to possess the properties stated in the proposition.

Proof of Property 1. Any permutation σ is a composition of transpositions, where a transposition is a permutation that only exchanges the positions of two elements. Let $\mathcal{H}(\mathbf{n}, d) = \{\mathbf{x} : \mathbf{n}^\mathsf{T}\mathbf{x} = d\}$ be a hyperplane with normal vector $\mathbf{n}$ and displacement d. For an ellipsoid centered at the origin, a transposition $\tau_{i,j}$ that exchanges axes i, j geometrically corresponds to a reflection w.r.t. the hyperplane $\mathcal{H}(\mathbf{e}_i - \mathbf{e}_j, 0) = \{\mathbf{x} : x_i = x_j\}$. Any reflection w.r.t. a hyperplane is an affine transformation $\mathbf{T}$. The center $\mathbf{c}$ of an invariant ellipsoid $\mathcal{E}$ must be fixed for any affine transformation $\mathbf{T}$ in its automorphism group $\mathrm{Aut}(\mathcal{E})$[24], i.e., $\mathbf{T}(\mathbf{c}) = \mathbf{c}$ for all $\mathbf{T} \in \mathrm{Aut}(\mathcal{E})$. In particular, $\mathbf{T}_{\tau_{i,j}}(\mathbf{c}) = \mathbf{c}$ for each $\mathbf{T}_{\tau_{i,j}}$ corresponding to reflection w.r.t. $\mathcal{H}(\mathbf{e}_i - \mathbf{e}_j, 0)$, i.e., for each transposition $\tau_{i,j}$. Equivalently, $\mathbf{c}$ is unchanged by reflection w.r.t. $\mathcal{H}(\mathbf{e}_i - \mathbf{e}_j, 0)$. Therefore $\mathbf{c} \in \mathcal{H}(\mathbf{e}_i - \mathbf{e}_j, 0)$, and thus $c_i = c_j$. As this is true for all i, j it follows that $c_1 = \cdots = c_n = c$, i.e., $\mathbf{c} = c\mathbf{1}$.

Proof of Property 2. We need to use the following facts and/or claims:

- A matrix $\mathbf{R}$ is diagonalizable (which is the case here since $\mathcal{E}$ is full-dimensional) if and only if all its eigenvectors span $\mathbb{R}^n$.
- The $\binom{n}{2}$ by n matrix with each row being the normal vector of a reflecting hyperplane $\mathcal{H}(\mathbf{e}_i - \mathbf{e}_j, 0)$ $(\forall i, j \in [n], i \neq j)$ for $\mathcal{E}$ (or Λ) has rank $n-1$. More concretely, these $n-1$ linearly

independent row vectors can be: $(\mathbf{e}_i - \mathbf{e}_{i+1})^\mathsf{T}$, $i \in [n-1]$. It is then easy to see how the rest rows of the matrix can be constructed from linear combination (just summation suffices) of these $n-1$ rows.

- Lemma 6 says that these normal vectors $(\mathbf{e}_i - \mathbf{e}_{i+1})$, $i \in [n-1]$ can be treated as eigenvectors of $\mathbf{R}$.

- Since these eigenvectors can span a subspace (or many subspaces) whose dimension (sum) is at most $n-1$, there must exist another eigenvector which is orthogonal to *all* these eigenvectors. Because $\mathbf{1}$ is the unique (up to normalization) vector orthogonal to all of them, $\mathbf{1}$ has to be one eigenvector of $\mathbf{R}$, meaning one axis of the ellipsoid is along the ray with direction $\mathbf{1}$ and associated semi-axis length a_1.

Proof of Property 3. Denote λ_{i+1} to be the eigenvalue associated with eigenvector $(\mathbf{e}_i - \mathbf{e}_{i+1})$ for $i \in [n-1]$. Because $(\mathbf{e}_i - \mathbf{e}_{i+1}) \not\perp (\mathbf{e}_{i+1} - \mathbf{e}_{i+2})$, it follows from Lem. 7 that $\lambda_{i+1} = \lambda_{i+2}\ \forall i \in [n-2]$. This means there actually exists a single eigen-subspace of dimension $n-1$ (Lem. 8), hence regardless of how the $n-1$ eigenvectors in this subspace are chosen, the ellipsoid's semi-axis lengths along these $n-1$ directions are all equal: $a_2 = \cdots = a_n$. □

Lem. 9 below gives an explicit construction for $\mathbf{R}^{-1}$, and its Cor. 2 allows us to characterize the ellipsoid that passes through $\{\mathbf{e}_i\}_{i=1}^n$. Finally, Lem. 10 gives an expression that must be satisfied in order for $\partial\mathcal{E}$ and $\partial\Lambda$ to share a common tangent point.

Lemma 9. *For any ellipsoid in the form of* (2.81)*, if* $\mathbf{q}_1 = \frac{1}{\sqrt{n}}\mathbf{1}$ *and* $\mathbf{D} = \mathrm{diag}(a_1^2, a_2^2, \ldots, a_2^2)$*, then* $\mathbf{R}^{-1} = \zeta\mathbf{1}_{n\times n} + a_2^{-2}\mathbf{I}_{n\times n}$*, where* $\zeta \equiv \frac{1}{n}\left(a_1^{-2} - a_2^{-2}\right)$*,* $\mathbf{1}_{n\times n}$ *is an* $n \times n$ *matrix with each element being* 1 *and* $\mathbf{I}_{n\times n}$ *is the* $n \times n$ *identity matrix.*

The proof of Lem. 9 is straightforward and omitted.

Corollary 2. *For any* $c > 1/n$ *and* $a_1 > \sqrt{n}(c - 1/n)$*, setting* a_2^2 *as below ensures* $\mathbf{e}_i \in \partial\mathcal{E}$ *for* $i \in [n]$

$$a_2^2 = \frac{(n-1)a_1^2}{na_1^2 - (nc-1)^2}. \tag{2.84}$$

Lemma 10. *Define* $\bar{\mathbf{p}}_t \equiv \mathbf{1} - \mathbf{p}_t$. *Then* $\partial\mathcal{E}$ *and* $\partial\Lambda$ *share a point of tangency at* $\mathbf{x}_t = \mathbf{x}(\mathbf{p}_t)$ *if for each* $i \in [n-1]$*:*

$$\left(\frac{a_2}{a_1}\right)^2 = \frac{(\bar{p}_{t,i} - \bar{p}_{t,n})\Sigma_{j=1}^n x_{t,j} + n(\bar{p}_{t,n}x_{t,i} - \bar{p}_{t,i}x_{t,n})}{(\bar{p}_{t,i} - \bar{p}_{t,n})(\Sigma_{j=1}^n x_{t,j} - nc)}. \tag{2.85}$$

Proof. Recall the bijection between $\mathbf{p}_t \in \partial\mathcal{S}$ and $\mathbf{x}_t \in \partial\Lambda$ (Cor. 1) established in[16]. Recall also Post[17] established that the tangent hyperplane to Λ at a point $\mathbf{x}_t \in \partial\Lambda$ with $\mathbf{x}_t \mapsto \mathbf{p}_t$ is $\left\{\mathbf{x} : (\mathbf{1} - \mathbf{p}_t)^\mathsf{T}\mathbf{x} = \prod_i (1 - p_{t,i})\right\}$. Next, using the implicit function theorem it is straightforward to establish that the tangent hyperplane to an arbitrary ellipsoid $\mathcal{E}$ (2.81) at a point $\mathbf{x}_t \in \partial\mathcal{E}$ is given by $\left\{\mathbf{x} : \mathbf{w}_t^\mathsf{T}\mathbf{x} = b\right\}$, where

$$\mathbf{w}_t = -\frac{\mathbf{R}^{-1}(\mathbf{x}_t - \mathbf{c}_0)}{\mathbf{R}_n^{-1}(\mathbf{x}_t - \mathbf{c}_0)},\ b = \mathbf{w}_t^\mathsf{T}\mathbf{x}_t. \tag{2.86}$$

Here $\mathbf{R}_n^{-1}$ is the n^{th} row of $\mathbf{R}^{-1}$. Finally equating (after appropriate normalization) these two tangent hyperplane equations yields the lemma. □

Because of the symmetries present in $\mathcal{E}(c, a_1, a_2)$, if we use a hyperplane with normal vector $\mathbf{1}$ to "slice" $\mathcal{E}(c, a_1, a_2)$ we get an $(n-1)$-dimensional ball. This is particularly intuitive in light of the following Prop. 12 for which we need to introduce the concept of rigid rotation of coordinate system.

Definition 9 (rigid rotation of coordinate system)**.** *Denote the original coordinate system as* $\mathbf{X}$*. A rigid rotation of the coordinate system about the origin is specified by two vectors* $\mathbf{v}_s$ *and* $\mathbf{v}_t$ *such that after the rotation* $\mathbf{v}_s$ *overlaps with* $\mathbf{v}_t$ *while during the rotation the relative position of* $\mathbf{v}_s$ *w.r.t. the coordinate axes remains unchanged. Denote this rotated coordinate system as* $\mathbf{U}$*, in which* $\mathbf{v}_s$ *was denoted (before rotation) as* $\mathbf{v}_t$ *in the original system* $\mathbf{X}$*. Alternatively, a rigid rotation is specified by a rotation matrix* $\mathbf{M}_{\text{rot}}$ *(that can be determined by the starting and terminating vectors* $\mathbf{v}_s, \mathbf{v}_t$[25]*) satisfying* $\mathbf{v}_t = \mathbf{M}_{\text{rot}}\mathbf{v}_s$*. All rotation matrices are orthonormal.*

According to this definition, we perform a rigid rotation of the original coordinate system $\mathbf{X}$ such that $\mathbf{e}_1$ overlaps with $\mathbf{1}/\sqrt{n}$ (i.e., $\mathbf{v}_s = \mathbf{e}_1$, $\mathbf{v}_t = \mathbf{1}/\sqrt{n}$), we then shift the origin to $\mathbf{c} = c\mathbf{1}$. Denote this rotated and translated system as $\mathbf{U}$, then the two systems are related by $\mathbf{X} = \mathbf{Q}\mathbf{U} + \mathbf{c}$ where $\mathbf{Q}$ is the associated rotation matrix.

Proposition 12. *In the above rotated and translated coordinate system* $\mathbf{U}$*,* Λ *has permutation symmetry among coordinates* $\{2,\ldots,n\}$*. If we let* $\mathbf{T}_u(\cdot)$ *denote this transformation from* $\mathbf{X}$ *to* $\mathbf{U}$*, namely* $\mathbf{u}=\mathbf{T}_u(\mathbf{x})=\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{c})$*, then:*

$$\mathbf{u}=(u_1,u_2,\ldots,u_n)^{\mathsf{T}}\in\mathbf{T}_u(\Lambda)\iff\sigma(\mathbf{u})=(u_1,u_{\sigma(2)},\ldots,u_{\sigma(n)})^{\mathsf{T}}\in\mathbf{T}_u(\Lambda),\ \forall\sigma\in S_{[n]\setminus\{1\}},\quad(2.87)$$

where $S_{[n]\setminus\{1\}}$ *is the permutation group whose members permute coordinate indices from* $\{2,\ldots,n\}$ *arbitrarily.*

Proof. Denote by S_A the permutation group of set A. Each member of this group can be viewed as a bijection. We overload the notation $\sigma(\cdot)$. First, if the input parameter is an n-vector it outputs the vector after permuting its coordinate components according to σ. Second, if the parameter is a natural number $k\in\{2,\ldots,n\}$ it only indicates how the k^{th} component is permuted, i.e., it gives the index of the coordinate taking the k^{th} position after permutation (assuming the original indexing is in natural order: $2,3,\cdots$). According to this notation, for any vector $\mathbf{v}=(v_1,v_2,\ldots,v_n)^{\mathsf{T}}$, we write $\sigma(\mathbf{v})=(v_1,v_{\sigma(2)},\ldots,v_{\sigma(n)})^{\mathsf{T}}$. Our goal is to show (2.87) namely $\mathbf{T}_u(\mathbf{x})\in\mathbf{T}_u(\Lambda)\iff\sigma(\mathbf{T}_u(\mathbf{x}))\in\mathbf{T}_u(\Lambda)$, $\forall\sigma\in S_{[n]\setminus\{1\}}$.

What we know can be seen from the following

$$\mathbf{T}_u(\mathbf{x})\in\mathbf{T}_u(\Lambda)\overset{(b)}{\iff}\mathbf{x}\in\Lambda\overset{(a)}{\iff}\sigma(\mathbf{x})\in\Lambda\overset{(b)}{\iff}\mathbf{T}_u(\sigma(\mathbf{x}))\in\mathbf{T}_u(\Lambda),\quad(2.88)$$

where (a) holds because of the natural inclusion for the symmetric group $S_{[n]}\supseteq S_{[n]\setminus\{1\}}$, and (b) holds because affine transformation preserves set membership. It therefore suffices to show

$$\mathbf{T}_u(\sigma(\mathbf{x}))=\sigma(\mathbf{T}_u(\mathbf{x})).\quad(2.89)$$

Recall $\mathbf{T}_u(\mathbf{x})=\mathbf{u}(\mathbf{x})=(u_1(\mathbf{x}),u_2(\mathbf{x}),\ldots,u_n(\mathbf{x}))^{\mathsf{T}}$ where we now highlight each component of $\mathbf{u}$ is

a function of the entire vector $\mathbf{x}$. We have

$$\mathbf{T}_u\left(\sigma(\mathbf{x})\right) = \left(u_1(\sigma(\mathbf{x})), u_2(\sigma(\mathbf{x})), \ldots, u_n(\sigma(\mathbf{x}))\right)^{\mathsf{T}} \tag{2.90}$$

$$\sigma(\mathbf{T}_u(\mathbf{x})) = \sigma\left(u_1(\mathbf{x}), u_2(\mathbf{x}), \ldots, u_n(\mathbf{x})\right)^{\mathsf{T}} = \left(u_1(\mathbf{x}), u_{\sigma(2)}(\mathbf{x}), \ldots, u_{\sigma(n)}(\mathbf{x})\right)^{\mathsf{T}}. \tag{2.91}$$

We need to show that

$$u_1(\sigma(\mathbf{x})) = u_1(\mathbf{x}), \quad \text{and} \quad u_j(\sigma(\mathbf{x})) = u_{\sigma(j)}(\mathbf{x}),\ \forall j \in [n] \setminus \{1\} \tag{2.92}$$

for all $\sigma \in S_{[n]\setminus\{1\}}$. Observe that obeying the symmetry condition (2.92) depends on the rotation and shift transformation $\mathbf{T}_u$. Define a matrix $\mathbf{Q}$ to be:

$$\mathbf{Q}_{ij} = \begin{cases} \frac{1}{\sqrt{n}}, & i = 1, \ldots, n,\ j = 1 \\ -\frac{1}{\sqrt{n}}, & i = 1,\ j = 2, \ldots, n \\ -\frac{1}{\sqrt{n}} + \frac{1}{1+\sqrt{n}}, & i \neq j,\ i = 2, \ldots, n,\ j = 2, \ldots, n \\ 1 - \frac{1}{\sqrt{n}} + \frac{1}{1+\sqrt{n}}, & i = j,\ i = 2, \ldots, n. \end{cases} \tag{2.93}$$

Namely, $\mathbf{Q}$ is the rotation matrix associated with the rigid rotation such that $\mathbf{e}_1$ becomes $\mathbf{1}/\sqrt{n}$ (Mortari[25] Eq. (10) pp. 7).

Therefore the two coordinate systems are related by $\mathbf{X} = \mathbf{QU} + \mathbf{c}$. Being a rotation matrix, $\mathbf{Q}^{-1} = \mathbf{Q}^{\mathsf{T}}$, so $\mathbf{Q}^{-1}_{i,j} = \mathbf{Q}^{\mathsf{T}}_{i,j} = \mathbf{Q}_{j,i}$. For any given point $\mathbf{x}$, $\mathbf{u} = \mathbf{Q}^{\mathsf{T}}(\mathbf{x} - \mathbf{c})$ and its components are:

$$u_1(\mathbf{x}) = \sum_{i=1}^{n} \mathbf{Q}^{\mathsf{T}}_{1,i}(\mathbf{x} - \mathbf{c})_i = \sum_{i=1}^{n} \mathbf{Q}_{i,1}(\mathbf{x} - \mathbf{c})_i = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_i - c) \tag{2.94}$$

$$\begin{aligned} u_j(\mathbf{x}) &= \sum_{i=1}^{n} \mathbf{Q}^{\mathsf{T}}_{j,i}(\mathbf{x} - \mathbf{c})_i = \mathbf{Q}^{\mathsf{T}}_{j,1}(\mathbf{x} - \mathbf{c})_1 + \mathbf{Q}^{\mathsf{T}}_{j,j}(\mathbf{x} - \mathbf{c})_j + \sum_{i=2, i\neq j}^{n} \mathbf{Q}^{\mathsf{T}}_{j,i}(\mathbf{x} - \mathbf{c})_i \\ &= \mathbf{Q}_{1,j}(\mathbf{x} - \mathbf{c})_1 + \mathbf{Q}_{j,j}(\mathbf{x} - \mathbf{c})_j + \sum_{i=2, i\neq j}^{n} \mathbf{Q}_{i,j}(\mathbf{x} - \mathbf{c})_i \\ &= -\frac{1}{\sqrt{n}}(x_1 - c) + \left(1 - \frac{1}{\sqrt{n}} + \frac{1}{1+\sqrt{n}}\right)(x_j - c) + \left(-\frac{1}{\sqrt{n}} + \frac{1}{1+\sqrt{n}}\right) \sum_{i=2, i\neq j}^{n} (x_i - c) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (x_i - c) + \frac{1}{1+\sqrt{n}} \sum_{i=2}^{n} (x_i - c) + (x_j - c), \qquad \forall j \in [n] \setminus \{1\}. \end{aligned} \tag{2.95}$$

Condition (2.92) can now be verified to be true. Hence we have shown (2.89) and this completes the proof. □

The previous proposition seems to suggest some loss of permutation symmetry. This is answered in the negative by the following.

Proposition 13 (conservation of symmetry)**.** *The permutation symmetry of the set* Λ *is preserved under rotation of the coordinate system.*

Proof. Let $S_{[n]}$ be the symmetric group of cardinality n, meaning its members permute coordinate indices from $[n]$ arbitrarily. A set G has permutation symmetry for coordinate indices $\mathcal{A} \subseteq [n]$ if G has permutation symmetry for $\Pi \subseteq S_{[n]}$ where $\Pi = \{\sigma : \sigma(k) = k,\ \forall k \notin \mathcal{A}\}$ (namely fixing the indices in $\mathcal{A}^c \equiv [n] \setminus \mathcal{A}$), meaning $x \in G \iff \sigma(x) \in G,\ \forall \sigma \in \Pi$. Any permutation can be decomposed as the product of a sequence of (disjoint) transpositions, where each transposition swaps two indices and keeps the remaining indices fixed. Since G is symmetric w.r.t. transposition of a pair of indices (i, j), this implies G has reflective symmetry w.r.t. the hyperplane $\mathcal{H}(\mathbf{n}_{i,j}, 0)$ for $\mathbf{n}_{i,j} = \mathbf{e}_i - \mathbf{e}_j$. It follows that if G has permutation symmetry for k out of n coordinates, then it has reflective symmetry w.r.t. each of the $\binom{k}{2}$ reflecting hyperplanes.

Now consider our particular transformation $\mathbf{T}_u$, as defined in Prop. 12. In the transformed space $\mathbf{U}$, $\mathbf{T}_u(\Lambda)$ has permutation symmetry w.r.t. coordinates $[n] \setminus \{1\}$, which gives $\binom{n-1}{2}$ degrees of reflective symmetry. Furthermore, $\mathbf{T}_u(\Lambda)$ retains reflective symmetry w.r.t. $\mathbf{T}_u(\mathcal{H}(\mathbf{n}_{1,j}, 0))$ for $j \in [n] \setminus \{1\}$ and this adds another $\binom{n-1}{1}$ degrees of symmetry. The total is $\binom{n-1}{2} + \binom{n-1}{1} \equiv \binom{n}{2}$, which agrees with the fact that Λ, in the original space $\mathbf{X}$, has full permutation symmetry (hence $\binom{n}{2}$ reflecting hyperplanes). □

The following two propositions comprise the second major contribution in this subsection. They justify why, for both $\mathcal{E}_{\text{out}}$ (Prop. 14) and $\mathcal{E}_{\text{in}}$ (Prop. 15), it suffices (necessity is clear) to only consider $\partial\Lambda$ in verifying an ellipsoid induces a valid inner or outer bound on Λ. This observation reduces the set of points that must be checked from $\mathbf{x} \in \Lambda$ to $\mathbf{x} \in \partial\Lambda$. More importantly, the characterization of $\partial\Lambda$ is amenable to analysis using majorization inequalities, which enables us to provide an alternative

proof of the proposed ellipsoid outer bound in §2.6.3.

Proposition 14. *Fix* $c > \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}$. *Then* $\Lambda \subseteq \Lambda_{\text{eo}} \iff \partial\Lambda \cap \mathcal{E}_{\text{out}} = \emptyset$.

Proof. "⇒": Because $\Lambda \subseteq \mathcal{S}$, so $\Lambda \subseteq \Lambda_{\text{eo}} = \mathcal{S} \setminus \mathcal{E}_{\text{out}}$ means $\Lambda \cap \mathcal{E}_{\text{out}} = \emptyset$, hence $\partial\Lambda \cap \mathcal{E}_{\text{out}} = \emptyset$.

"⇐" (We prove the contrapositive version namely $\Lambda \not\subseteq \Lambda_{\text{eo}} \Rightarrow \partial\Lambda \cap \mathcal{E}_{\text{out}} \neq \emptyset$):

Suppose $\Lambda \not\subseteq \Lambda_{\text{eo}}$, then $\Lambda \cap \mathcal{E}_{\text{out}} \neq \emptyset$, meaning there exists $\mathbf{x}_0 \in \Lambda \cap \mathcal{E}_{\text{out}}$. If $\mathbf{x}_0 \in \partial\Lambda$, this already says $\partial\Lambda \cap \mathcal{E}_{\text{out}} \neq \emptyset$. If $\mathbf{x}_0 \notin \partial\Lambda$, due to the convexity of $\mathcal{E}_{\text{out}}$, line segment $\overline{\mathbf{x}_0\mathbf{c}} \in \mathcal{E}_{\text{out}}$, where $\mathbf{c}$ is the center of $\mathcal{E}_{\text{out}}$. Since $\mathbf{c} = c\mathbf{1}$ where $c > \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}$, so $\mathbf{c} \notin \Lambda$. But because $\mathbf{x}_0 \in \Lambda$, the line segment $\overline{\mathbf{x}_0\mathbf{c}}$ must necessarily cross $\partial\Lambda$. □

Proposition 15. *Fix* $c > \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}$. *Then* $\Lambda_{\text{ei}} \subseteq \Lambda \iff \partial\Lambda \subseteq \overline{\mathcal{E}_{\text{in}}}$.

Proof. This proof (converse part) directly invokes Lem. 11 and 15 where the proof of Lem. 15 uses Lem. 11, 12 , 13 and 14. All these five lemmas are shown below within the proof of this proposition. In this proof, all set complements are w.r.t. the simplex $\mathcal{S}$. Recall $\Lambda \subseteq \mathcal{S}$.

"⇒":

$$\Lambda_{\text{ei}} \subseteq \Lambda \quad \Longleftrightarrow \quad \Lambda^c \subseteq \Lambda_{\text{ei}}^c = (\mathcal{S} \setminus \mathcal{E}_{\text{in}})^c = \mathcal{E}_{\text{in}} \cap \mathcal{S} \subseteq \overline{\mathcal{E}_{\text{in}}} \tag{2.96}$$

$$\overset{(a)}{\Rightarrow} \quad \overline{\Lambda^c} \subseteq \overline{\mathcal{E}_{\text{in}}} \quad \overset{(b)}{\Rightarrow} \quad \partial\Lambda \subseteq \overline{\mathcal{E}_{\text{in}}}, \tag{2.97}$$

where (a) is because for all sets A, B we have $A \subseteq \overline{B} \Rightarrow \overline{A} \subseteq \overline{B}$, and (b) is due to the definition $\text{bd}(A) = \overline{A} \cap \overline{A^c}$.

"⇐" : Given $\partial\Lambda \subseteq \overline{\mathcal{E}_{\text{in}}}$, we need to show $\forall \mathbf{x}_0 \in \Lambda_{\text{ei}} \equiv \mathcal{S} \setminus \mathcal{E}_{\text{in}}$, it holds true that $\mathbf{x}_0 \in \Lambda$. Note $\mathbf{x}_0 \in \mathcal{S}$ and because $\mathcal{S}$ can be written as the disjoint union $\Lambda \cup (\mathcal{S} \cap \Lambda^c)$, so equivalently we need to show $\mathbf{x}_0 \in \mathcal{S} \cap \Lambda^c$ is impossible. We prove by contradiction. Assuming $\mathbf{x}_0 \in \mathcal{S} \cap \Lambda^c$, because $\overline{\Lambda^c} \cap \mathcal{S} = \text{conv}(\partial\Lambda)$ (Lem. 11) and the fact $\text{bd}(A^c) = \text{bd}(A)$, we have (where the last step follows from Lem. 15)

$$\mathbf{x}_0 \in \left(\overline{\Lambda^c} \cap \mathcal{S}\right) \setminus \text{bd}(\Lambda^c) = \text{conv}(\partial\Lambda) \setminus \partial\Lambda \subseteq \mathcal{E}_{\text{in}}. \tag{2.98}$$

But this contradicts the assumption $\mathbf{x}_0 \in \mathcal{S} \setminus \mathcal{E}_{\text{in}}$.

We now present the five lemmas mentioned above.

Lemma 11. $\overline{\Lambda^c} \cap \mathcal{S} = \text{conv}(\partial\Lambda)$.

Proof of Lem. 11: Since $\partial\Lambda$ is also the boundary of the closed set $\overline{\Lambda^c}$, we have: $\partial\Lambda \subseteq \overline{\Lambda^c}$. Because $\partial\Lambda \subseteq \Lambda \subseteq \mathcal{S}$, we have $\partial\Lambda \subseteq \overline{\Lambda^c} \cap \mathcal{S}$, and hence $\text{conv}(\partial\Lambda) \subseteq \overline{\Lambda^c} \cap \mathcal{S}$ due to convexity. The proof will be complete if we show $\overline{\Lambda^c} \cap \mathcal{S} \subseteq \text{conv}(\partial\Lambda)$. Given $\mathbf{x} \in \overline{\Lambda^c} \cap \mathcal{S}$, we distinguish cases: *i*) if $\mathbf{x} \in \text{bd}(\Lambda) = \partial\Lambda$, done; *ii*) if $\mathbf{x} \in \partial\mathcal{S}$, since $\partial\mathcal{S} = \text{conv}(\mathbf{e}_1, \ldots, \mathbf{e}_n) \subseteq \text{conv}(\partial\Lambda)$, it follows $\mathbf{x} \in \text{conv}(\partial\Lambda)$; *iii*) otherwise we can draw a line along the direction $\mathbf{1}$ that passes $\mathbf{x}$, since (in this case) this line must necessarily intersect at distinct points on $\partial\mathcal{S}$ and $\partial\Lambda$, this shows $\mathbf{x} \in \text{conv}(\partial\Lambda)$ again. □

Lemma 12. *Fix two closed sets* $\overline{A}$, $\overline{B}$. *Then* $\overline{A} \subseteq \overline{B} \Rightarrow \overline{A} \setminus \text{bd}(\overline{A}) \subseteq \overline{B} \setminus \text{bd}(\overline{B})$.

Proof of Lem. 12: Equivalently, we show $\overline{A} \subseteq \overline{B} \Rightarrow \text{int}(\overline{A}) \subseteq \text{int}(\overline{B})$. Given $\mathbf{x}_0 \in \text{int}(\overline{A})$ meaning there exists an open ball $\mathcal{B}(\mathbf{x}_0, \epsilon) = \{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\| < \epsilon\} \subseteq \overline{A} \subseteq \overline{B}$, this then directly says $\mathbf{x}_0 \in \text{int}(\overline{B})$. □

Lemma 13. *Fix sets* A, B, C *satisfying* $C \subseteq A$. *Then* $A \setminus B = (A \setminus (B \cup C)) \cup (C \setminus B)$.

Proof of Lem. 13: "$\subseteq$": Given $\mathbf{x}_0 \in A \setminus B$, there are two cases: *i*) $\mathbf{x}_0 \notin C$, and hence $\mathbf{x}_0 \in A, \notin B \cup C$ namely $\mathbf{x}_0 \in A \setminus (B \cup C)$; *ii*) $\mathbf{x}_0 \in C$, and hence $\mathbf{x}_0 \in A, \in C, \notin B$ namely $\mathbf{x}_0 \in A \cap (C \setminus B) = C \setminus B$ because $C \subseteq A$. "$\supseteq$": Given $\mathbf{x}_0 \in (A \setminus (B \cup C)) \cup (C \setminus B)$, either $\mathbf{x}_0 \in A \setminus (B \cup C)$ or $\mathbf{x}_0 \in C \setminus B$ gives $\mathbf{x}_0 \in A \setminus B$, as $C \subseteq A$. □

Lemma 14. *If* $\partial\mathcal{S} \subseteq \overline{\mathcal{E}_{\text{in}}}$, *then* $\partial\mathcal{S} \cap \partial\mathcal{E}_{\text{in}} \subseteq \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$.

Proof of Lem. 14: Suppose $\partial\mathcal{S} \cap \partial\mathcal{E}_{\text{in}} \not\subseteq \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$, then $\exists \mathbf{y} \in \partial\mathcal{S} \cap \partial\mathcal{E}_{\text{in}}$ that is not an extreme point (i.e., $\mathbf{e}_i$'s) of $\partial\mathcal{S}$, meaning $\mathbf{y}$ can be represented as a convex combination of two distinct points $\mathbf{y}_1, \mathbf{y}_2 \in \partial\mathcal{S}$. As $\partial\mathcal{S} \subseteq \overline{\mathcal{E}_{\text{in}}}$, this then implies $\mathbf{y}$ is not an extreme point of $\overline{\mathcal{E}_{\text{in}}}$ either. However, it is clear that all points on $\partial\mathcal{E}_{\text{in}}$ (including $\mathbf{y}$) are extreme points of $\overline{\mathcal{E}_{\text{in}}}$, contradiction. □

Lemma 15. *If* $\partial\Lambda \subseteq \overline{\mathcal{E}_{\text{in}}}$, *then* $\text{conv}(\partial\Lambda) \setminus \partial\Lambda \subseteq \mathcal{E}_{\text{in}}$.

Proof of Lem. 15: Given $\partial\Lambda \subseteq \overline{\mathcal{E}_{\text{in}}}$, it follows $\text{conv}(\partial\Lambda) \subseteq \overline{\mathcal{E}_{\text{in}}}$, namely (Lem. 11) $\overline{\Lambda^c} \cap \mathcal{S} \subseteq \overline{\mathcal{E}_{\text{in}}}$. As $\text{bd}(\text{conv}(\partial\Lambda)) = \partial\Lambda \cup \partial\mathcal{S}$, Lem. 12 then gives $\text{conv}(\partial\Lambda) \setminus (\partial\Lambda \cup \partial\mathcal{S}) \subseteq \mathcal{E}_{\text{in}}$. Applying Lem. 13 with $A = \text{conv}(\partial\Lambda)$, $B = \partial\Lambda$ and $C = \partial\mathcal{S}$, we have $\text{conv}(\partial\Lambda) \setminus \partial\Lambda = (\text{conv}(\partial\Lambda) \setminus (\partial\Lambda \cup \partial\mathcal{S})) \cup (\partial\mathcal{S} \setminus \partial\Lambda)$, so to finish the proof we need to show $\partial\mathcal{S} \setminus \partial\Lambda \subseteq \mathcal{E}_{\text{in}}$. Towards this, recall $\partial\mathcal{S} \subseteq \text{conv}(\partial\Lambda) \subseteq \overline{\mathcal{E}_{\text{in}}}$, for a point $\mathbf{x} \in \partial\mathcal{S} \subseteq \overline{\mathcal{E}_{\text{in}}}$, either $\mathbf{x} \notin \partial\mathcal{E}_{\text{in}}$ or $\mathbf{x} \in \partial\mathcal{E}_{\text{in}}$. If $\mathbf{x} \notin \partial\mathcal{E}_{\text{in}}$ this means $\mathbf{x} \in \mathcal{E}_{\text{in}}$. If $\mathbf{x} \in \partial\mathcal{E}_{\text{in}}$ then $\mathbf{x} \in \partial\mathcal{S} \cap \partial\mathcal{E}_{\text{in}}$ and hence (Lem. 14) $\mathbf{x} \in \{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ which means $\mathbf{x} \in \partial\Lambda$. So we have proved $\partial\mathcal{S} \setminus \partial\Lambda \subseteq \mathcal{E}_{\text{in}}$. □

This concludes the proof of Prop. 15. □

The following proposition concludes this subsection; it further reduces the search space of points that must be checked to establish the correctness of a bound on Λ induced by an ellipsoid $\mathcal{E}(c, a_1, a_2)$. Define $\mathcal{V}(\mathbf{p}) = \{a \in (0, 1] : \exists i \in [n] : p_i = a\}$ as the set of non-zero values taken by a $\mathbf{p} \in \partial\mathcal{S}$. Further define $\mathcal{P}^{(2)} = \{\mathbf{p} \in \partial\mathcal{S} : |\mathcal{V}(\mathbf{p})| \in \{1, 2\}\}$ as those $\mathbf{p}$ taking at most two distinct non-zero values, and $\mathcal{P}^{(1)} = \{\mathbf{p} \in \partial\mathcal{S} : |\mathcal{V}(\mathbf{p})| = 1\}$ as those $\mathbf{p}$ taking exactly one non-zero value; we refer to $\mathcal{P}^{(1)}$ as the set of quasi-uniform (QU) vectors. The following proposition reduces the search space from $\partial\Lambda$ to $\{\mathbf{x}(\mathbf{p}) : \mathbf{p} \in \mathcal{P}^{(2)}\}$.

Proposition 16. *Fix an ellipsoid $\mathcal{E}(c, a_1, a_2)$. A potential extremizer of the maximization or minimization problem*

$$\max_{\mathbf{p} \in \partial\mathcal{S}} \ (\text{or} \min_{\mathbf{p} \in \partial\mathcal{S}}) \ f(\mathbf{x}(\mathbf{p})) \equiv (\mathbf{x}(\mathbf{p}) - \mathbf{c})^{\mathsf{T}} \mathbf{R}^{-1} (\mathbf{x}(\mathbf{p}) - \mathbf{c}) \tag{2.99}$$

can have at most two distinct values among all its non-zero component(s), i.e., $\mathbf{p} \in \mathcal{P}^{(2)}$.

Proof. For an ellipsoid in the form of (2.81), membership testing is by comparing the objective function $f(\mathbf{x}(\mathbf{p}))$ (2.99) with 1. Restricting the feasible set in (2.99) to $\mathbf{p} \in \partial\mathcal{S}$ follows from the results in §2.6.1 (Props. 14 and 15). Note maximization and minimization is for inner and outer bounding ellipsoid, respectively. Introducing Lagrange multipliers μ, $(\lambda_i, i \in [n])$ for the equality constraint and each inequality constraint respectively, the Lagrangian of this optimization problem is:

$$\mathcal{L}(\mathbf{p}, \mu, \boldsymbol{\lambda}) = f + \mu \left(\sum_{i=1}^{n} p_i - 1 \right) + \sum_{i=1}^{n} \lambda_i \left(-p_i \right). \tag{2.100}$$

The first-order Karush-Kuhn-Tucker (KKT) necessary conditions for a local extremizer are:

$$\begin{aligned}
\text{stationarity} \qquad & \frac{\partial \mathcal{L}}{\partial p_i} = 0,\ i \in [n] \\
\text{primal feasibility} \qquad & \sum_{i=1}^{n} p_i - 1 = 0;\ -p_i \leq 0,\ i \in [n] \\
\text{dual feasibility} \qquad & \lambda_i \leq (\text{or } \geq)\ 0,\ i \in [n] \\
\text{complementary slackness} \qquad & \lambda_i\,(-p_i) = 0,\ i \in [n].
\end{aligned}$$

Fix $\mathbf{p} \in \partial\mathcal{S}$ such that $|\mathcal{V}(\mathbf{p})| \geq 2$. As $\frac{\partial \mathcal{L}}{\partial p_i} = \frac{\partial f}{\partial p_i} + \mu - \lambda_i$, it follows that if a potential local extremizer $\mathbf{p}$ has two distinct non-zero components, say $0 < p_k < p_l < 1$, then, due to complementary slackness, the stationarity of the Lagrangian reduces to the equality of derivatives of the objective function w.r.t. p_k and p_l:

$$\frac{\partial \mathcal{L}}{\partial p_k} = \frac{\partial \mathcal{L}}{\partial p_l} = 0 \Leftrightarrow \frac{\partial f}{\partial p_k} - \frac{\partial f}{\partial p_l} = 0 \Leftrightarrow \frac{1}{2}\left(\frac{\partial f}{\partial p_k} - \frac{\partial f}{\partial p_l}\right)(1 - p_k)\,(1 - p_l) = 0. \tag{2.101}$$

Let $\mathbf{x} = \mathbf{x}(\mathbf{p})$. For notational convenience, we make the following definitions:

$$\begin{aligned}
&\Sigma \equiv \sum_i x_i, \quad S \equiv \sum_i x_i^2, \qquad \pi_i \equiv \prod_{j \neq i}(1 - p_j), \quad \pi_{ik} \equiv \prod_{j \neq i,k}(1 - p_j),\ \forall i, k \in [n] \qquad (2.102) \\
&\beta_i \equiv \frac{1}{n a_1^2}(\Sigma - nc) + \frac{1}{a_2^2}\left(x_i - \frac{1}{n}\Sigma\right),\ i \in [n]; \quad \alpha \equiv \frac{1}{n a_1^2}\left(-\Sigma^2 + nc\Sigma\right) + \frac{1}{a_2^2}\left(-S + \frac{1}{n}\Sigma^2\right).
\end{aligned}$$

Applying Lem. 9, we compute

$$f(\mathbf{x}(\mathbf{p})) = (\mathbf{x} - c\mathbf{1})^{\mathsf{T}}\left(\zeta \mathbf{1}_{n\times n} + a_2^{-2}\mathbf{I}_{n\times n}\right)(\mathbf{x} - c\mathbf{1}) = \frac{1}{n}(\Sigma - nc)^2 a_1^{-2} + \left(S - \frac{1}{n}\Sigma^2\right) a_2^{-2}. \tag{2.103}$$

We can also check

$$\begin{aligned}
\frac{\partial \Sigma}{\partial p_i} &= -\frac{1}{1-p_i}\Sigma + \frac{1}{p_i(1-p_i)}x_i && (2.104)\\
\frac{\partial S}{\partial p_i} &= 2\left(-\frac{1}{1-p_i}S + \frac{1}{p_i(1-p_i)}x_i^2\right) && (2.105)\\
\frac{\partial f}{\partial p_i} &= \frac{1}{na_1^2}2\left(\Sigma - nc\right)\frac{\partial \Sigma}{\partial p_i} + \frac{1}{a_2^2}\left(\frac{\partial S}{\partial p_i} - \frac{1}{n}2\Sigma\frac{\partial \Sigma}{\partial p_i}\right) && (2.106)\\
\frac{\partial f}{\partial p_i}\frac{1}{2}(1-p_i) &= \alpha + \pi_i\beta_i. && (2.107)
\end{aligned}$$

Using these expressions, the RHS of (2.101) may be written as

$$(1-p_l)(\alpha - (-\pi_k\beta_k)) = (1-p_k)(\alpha - (-\pi_l\beta_l)). \tag{2.108}$$

It can be shown that $\alpha = \sum_{i=1}^n p_i(-\pi_i\beta_i) = \mathbb{E}_{\mathbf{p}}(-\pi_i\beta_i)$, so the only way to satisfy equality (2.108) is to require $(-\pi_i\beta_i)$'s associated with every non-zero components of $\mathbf{p} \notin \mathcal{P}^{(1)}$ to be all equal, i.e., at this point we have established

$$\mathbf{p} : |\mathcal{V}(\mathbf{p})| \geq 2 \text{ and } \mathbf{p} \text{ satisfies (2.101)} \;\Rightarrow\; \pi_i(\mathbf{p})\beta_i(\mathbf{p}) \text{ are equal for all } i \in [n] : p_i > 0. \tag{2.109}$$

We next investigate the consequence of this property. First observe $\pi_i \neq 0$ always holds. There are two possibilities: *i*) there exists some $i \in [n]$ with $\beta_i = 0$, or *ii*) $\beta_i \neq 0$ for all $i \in [n]$. First consider case *i*). If some $\beta_i = 0$, then all β_i's must equal zero (as otherwise $\pi_i\beta_i = \pi_j\beta_j$ won't hold). It then follows from the definition of β_i that all the non-zero components of $\mathbf{p}$ are the same (since all the x_i's are the same), meaning $|\mathcal{V}(\mathbf{p})| = 1$, contradicting the assumption that $|\mathcal{V}(\mathbf{p})| \geq 2$.

It remains to consider case *ii*), i.e., $\pi_i \neq 0$ and $\beta_i \neq 0$ for all i such that $p_i \neq 0$. W.l.o.g., let i, j be indices associated with two non-zero components of $\mathbf{p}$. Manipulations of the expressions for π_i and β_i yield

$$\frac{\pi_i}{\pi_j} = \frac{\beta_j}{\beta_i} \text{ and } p_i \neq p_j \Rightarrow \frac{\pi_i}{\beta_j} = \frac{\pi_j}{\beta_i} = -a_2^2. \tag{2.110}$$

Since the LHS of the above equation holds for the assumed indices i, j because of (2.109). The RHS

of (2.110) therefore holds, and we can prove by contradiction that there are *at most two* distinct values among all the non-zero components of a potential local extremizer. Namely, if there are indices i, j, k with $0 < p_i < p_j < p_k < 1$, then applying the RHS of (2.110) to each of the three pairs of indices, i.e., $\{i, j\}, \{i, k\}, \{j, k\}$, will necessarily violate the assumption $0 < p_i < p_j < p_k < 1$. This completes the proof that a potential extremizer $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})| \geq 2$ must have $|\mathcal{V}(\mathbf{p})| = 2$, i.e., $\mathbf{p} \in \mathcal{P}^{(2)}$.

We have now established Prop. 16; the remainder of this proof will establish some expressions that will be of use in Props. 18 and 17. A consequence of the above proof is the following. Suppose $\mathbf{p}$ has $|\mathcal{V}(\mathbf{p})| = 2$ and let i, j be indices such that $0 < p_i < p_j < 1$. Then

$$\frac{nc - \Sigma}{a_1^2} = \frac{n(\pi_i + x_j) - \Sigma}{a_2^2} = \frac{n(\pi_j + x_i) - \Sigma}{a_2^2}. \tag{2.111}$$

Substitutions of the expressions in (2.102) and (2.103) to the objective function f in (2.103) yields

$$f(\mathbf{x}(\mathbf{p})) = -\alpha + \frac{c}{a_1^2}(nc - \Sigma). \tag{2.112}$$

If an extremizer $\mathbf{p}$ has $|\mathcal{V}(\mathbf{p})| = 2$ and satisfies (2.111), then the objective function may be expressed as

$$f(\mathbf{x}(\mathbf{p})) = \frac{1}{a_2^2}\left(-\pi_i \pi_j + cn\left(\pi_i + x_j - \frac{\Sigma}{n}\right)\right). \tag{2.113}$$

We emphasize however that this expression for f is derived under the assumptions and constraints associated with (2.110), and is not valid in general. □

2.6.2 Explicit ellipsoid induced bounds

The key contributions of this subsection are Props. 18 and 17 (see §2.8 for proofs), which leverage the results from the previous subsection to provide an explicit construction for ellipsoid induced inner and outer bounds on Λ. As with the proofs of the spherical inner and outer bounds, Λ_{si} and Λ_{so}, the key proof technique we use is by exploiting the implications of Karush-Kuhn-Tucker first order necessary conditions. In the next subsection (§2.6.3) we establish the ellipsoid outer bound

using a different proof technique.

Proposition 17 (ellipsoid outer bound)**.** *The ellipsoid* $\mathcal{E}_{\text{out}}(c, a_{1,\text{out}}(c), a_{2,\text{out}}(c))$ *(in the class of* $\mathcal{E}(c, a_1, a_2)$ *ellipsoids) with*

$$a_{1,\text{out}}(c) = \sqrt{(nc-1)c}, \quad a_{2,\text{out}}(c) = \sqrt{\frac{(n-1)}{na_{1,\text{out}}^2 - (nc-1)^2}}a_{1,\text{out}} = \sqrt{(n-1)c} \tag{2.114}$$

induces an outer bound $\Lambda_{\text{eo}} = \mathcal{S} \setminus \mathcal{E}_{\text{out}}$ *on* Λ*, for all* $c > 1/n$*.*

Proposition 18 (ellipsoid inner bound)**.** *The ellipsoid* $\mathcal{E}_{\text{in}}(c, a_{1,\text{in}}(c), a_{2,\text{in}}(c))$ *(in the class of* $\mathcal{E}(c, a_1, a_2)$ *ellipsoids) with*

$$a_{1,\text{in}}(c) = \sqrt{n}(c-m), \quad a_{2,\text{in}}(c) = \sqrt{\frac{(n-1)}{na_{1,\text{in}}^2 - (nc-1)^2}}a_{1,\text{in}} \tag{2.115}$$

induces an inner bound $\Lambda_{\text{ei}} = \mathcal{S} \setminus \mathcal{E}_{\text{in}}$ *on* Λ*, for all* $c > 1/n$*.*

Remark 5. *The ellipsoid inner bound* Λ_{ei} *recovers the optimal spherical inner bound* Λ_{si}^* *by setting* c *to be the critical* $c_{\text{in}}^* = (1-nm^2)/(2(1-nm)$ *(note* $c_{\text{in}}^* > 1/n$ *for all* $n \geq 2$*); the ellipsoid outer bound* Λ_{eo} *recovers the optimal spherical outer bound* Λ_{so}^* *by setting* $c = c_{\text{out}}^* = 1$*. Furthermore, as a consequence of the tightness monotonicity Prop. 19, any* Λ_{ei} *with* $c > c_{\text{in}}^*$ *is better than* Λ_{si}^* *and any* Λ_{eo} *with* $c > 1$ *is better than* Λ_{so}^* *(in terms of volume approximation to* Λ*).*

In words, $\mathcal{E}_{\text{in}}$ is such that its boundary $\partial\mathcal{E}_{\text{in}}$ passes through $\mathbf{e}_1, \ldots, \mathbf{e}_n$ and the all-rates-equal point $\mathbf{m}$. It can be shown that $\partial\mathcal{E}_{\text{in}}$ is tangent at $\mathbf{m}$ with $\partial\Lambda$ (recall Lem. 10). $\mathcal{E}_{\text{out}}$ is such that its boundary $\partial\mathcal{E}_{\text{out}}$ passes through each $\mathbf{e}_i$ and is further tangent with $\partial\Lambda$ at each $\mathbf{e}_i$ (recall Lem. 10).

The ellipsoid bounds Λ_{ei}, Λ_{eo} together with $\partial\Lambda$ are shown in Fig. 2.7 for $n = 2$ and 3, where the center is chosen to be $c = 2$. Improvements can be seen by comparing this figure with the optimal spherical bounds shown in Fig. 2.6. Furthermore, the quality of the ellipsoid bounds are improved by increasing c, as stated below.

Proposition 19. *For all* $c > 1/n$*, the tightness of both* Λ_{ei} *and* Λ_{eo} *increases in* c*.*

This result is illustrated in Fig. 2.8.

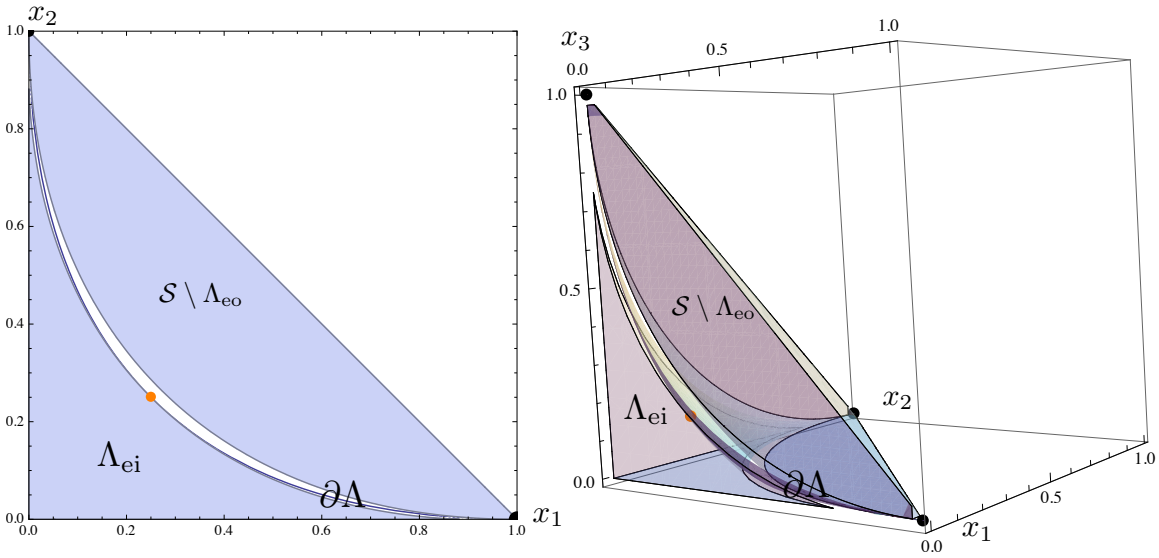

Figure 2.7: Ellipsoid bounds for $n = 2$ (left) and 3 (right) when $c = 2$: inner bound and the complement (w.r.t. $\mathcal{S}$) of outer bound are shown in solid; in between sits $\partial\Lambda$. Also shown are the all-rates-equal point $\mathbf{m}$ (orange) and corner points $\mathbf{e}_i$'s (black).

Proof. It will be easier to work in a rotated coordinate system, and further to consider all possible "cross-sections" of an ellipsoid $\mathcal{E}$, obtained by intersecting $\mathcal{E}$ with the hyperplane with normal vector $\mathbf{1}$. First, perform a rigid rotation of the original coordinate system $\mathbf{X}$ according to Def. 9 such that $\mathbf{e}_1$ overlaps with $\mathbf{1}/\sqrt{n}$ (as in Prop. 12). Second, shift the origin to the point $\mathbf{1}/n$, namely the intersection between $\partial\mathcal{S}$ and vector $\mathbf{1}$. Relabel this coordinate system as $\mathbf{X}$. Consequently, the original $\partial\mathcal{S}$ completely resides in the new coordinate hyperplane $\{\mathbf{x} : x_1 = 0\}$, and the equation for the boundary of ellipsoid $\mathcal{E}(c, a_1, a_2)$ has a standard form

$$\frac{\left(x_1 - \sqrt{n}\left(c - \frac{1}{n}\right)\right)^2}{a_1^2} + \frac{1}{a_2^2}\sum_{i=2}^{n} x_i^2 = 1. \tag{2.116}$$

By construction of the inner bound, establishing the proposition requires we show that if $c_1 < c_2$ then $\Lambda_{\text{ei}}(c_1) \subseteq \Lambda_{\text{ei}}(c_2)$, namely, if $c_1 < c_2$ then $\mathcal{E}_{\text{in}}(c_1) \cap \mathcal{S} \supseteq \mathcal{E}_{\text{in}}(c_2) \cap \mathcal{S}$. Likewise, by construction of the outer bound, establishing the proposition requires we show that if $c_1 < c_2$ then $\Lambda_{\text{eo}}(c_1) \supseteq \Lambda_{\text{eo}}(c_2)$, namely, if $c_1 < c_2$ then $\mathcal{E}_{\text{out}}(c_1) \cap \mathcal{S} \subseteq \mathcal{E}_{\text{out}}(c_2) \cap \mathcal{S}$.

Because the ellipsoid $\mathcal{E}(c, a_1, a_2)$ is spherical on the cross-section and for co-centered spheres

the relative largeness of radius is the sole indicator of inclusion relationship, we can work with the cross-section where the displacement is negative (corresponding to sitting inside $\mathcal{S}$, recall we rotated the system). From the standard form equation (2.116) we can solve for the square of the radius of the sphere (with height x_1) on the cross-section as

$$r_2^2 \equiv a_2^2\left(1-\frac{\left(x_1-\sqrt{n}\left(c-\frac{1}{n}\right)\right)^2}{a_1^2}\right)=\frac{n-1}{n}\frac{a_1^2-\left(x_1-\sqrt{n}\left(c-\frac{1}{n}\right)\right)^2}{a_1^2-\left(\sqrt{n}\left(c-\frac{1}{n}\right)\right)^2} \tag{2.117}$$

where we substitute the expressions from Props. 18 and 17.

Thus for Λ_{ei} we need to show for $\mathbf{x}\in\overline{\mathcal{E}_{\text{in}}}(c,a_1,a_2)$ with $x_1\le 0$, the above r_2 is decreasing in c; for Λ_{eo} we need to show for $\mathbf{x}\in\overline{\mathcal{E}_{\text{out}}}(c,a_1,a_2)$ with $x_1\le 0$, the above r_2 is increasing in c.

We first check Λ_{ei}. To ensure $\mathbf{x}\in\overline{\mathcal{E}_{\text{in}}}$, we must have $x_1\ge-\sqrt{n}\,(1/n-m)$. Substituting $a_{1,\text{in}}^2$ from (2.115) in Prop. 18 into (2.117), we have

$$r_{2,\text{in}}^2=\frac{n-1}{n}\frac{(2c-(m+y_{1,\text{in}}))\,(y_{1,\text{in}}-m)}{\left(2c-\left(m+\frac{1}{n}\right)\right)\left(\frac{1}{n}-m\right)},\quad\text{where } y_{1,\text{in}}\equiv\frac{x_1}{\sqrt{n}}+\frac{1}{n}\in\left[m,\frac{1}{n}\right]. \tag{2.118}$$

We take the derivative of $\frac{(2c-(m+y_{1,\text{in}}))}{\left(2c-\left(m+\frac{1}{n}\right)\right)}$ w.r.t. c and obtain $\frac{2\left(y_{1,\text{in}}-\frac{1}{n}\right)}{\left(2c-\left(m+\frac{1}{n}\right)\right)^2}\le 0$ for all $y_{1,I}\in\left[m,\frac{1}{n}\right]$. This shows $r_{2,\text{in}}$ is decreasing in c for the regimes of interest.

We next check Λ_{eo}. To ensure $\mathbf{x}\in\overline{\mathcal{E}_{\text{out}}}$, we must have $x_1\ge-\sqrt{n}\sqrt{c-\frac{1}{n}}\left(\sqrt{c}-\sqrt{c-\frac{1}{n}}\right)$ (this is obtained by solving for the intersecting point of $\partial\mathcal{E}_{\text{out}}$ with coordinate axis X_1). Substituting $a_{1,\text{out}}^2$ from (2.114) in Prop. 17 into (2.117), we have

$$r_{2,\text{out}}^2=(n-1)\left(c-\frac{(c-y_{1,\text{out}})^2}{c-\frac{1}{n}}\right),\quad\text{where } y_{1,\text{out}}\equiv\frac{x_1}{\sqrt{n}}+\frac{1}{n}\in\left[\sqrt{c}\left(\sqrt{c}-\sqrt{c-\frac{1}{n}}\right),\frac{1}{n}\right]. \tag{2.119}$$

We take the derivative of $\left(c-\frac{(c-y_{1,\text{out}})^2}{c-\frac{1}{n}}\right)$ w.r.t. c and obtain $1-\frac{c-y_{1,\text{out}}}{\left(c-\frac{1}{n}\right)^2}\left(2\left(c-\frac{1}{n}\right)-(c-y_{1,\text{out}})\right)$ which is always non-negative since its non-negativeness is equivalent to $(y_{1,\text{out}}-1/n)^2\ge 0$. This also implies we don't need to restrict our attention to $\mathcal{S}$ (corresponding to $y_{1,\text{out}}\le 1/n$). There is a complete set inclusion relationship as c increases. □

Thus $\partial\Lambda$ is increasingly tightly "sandwiched" between part of $\partial\mathcal{E}_{\text{in}}$ and $\partial\mathcal{E}_{\text{out}}$ as $c\to\infty$. This

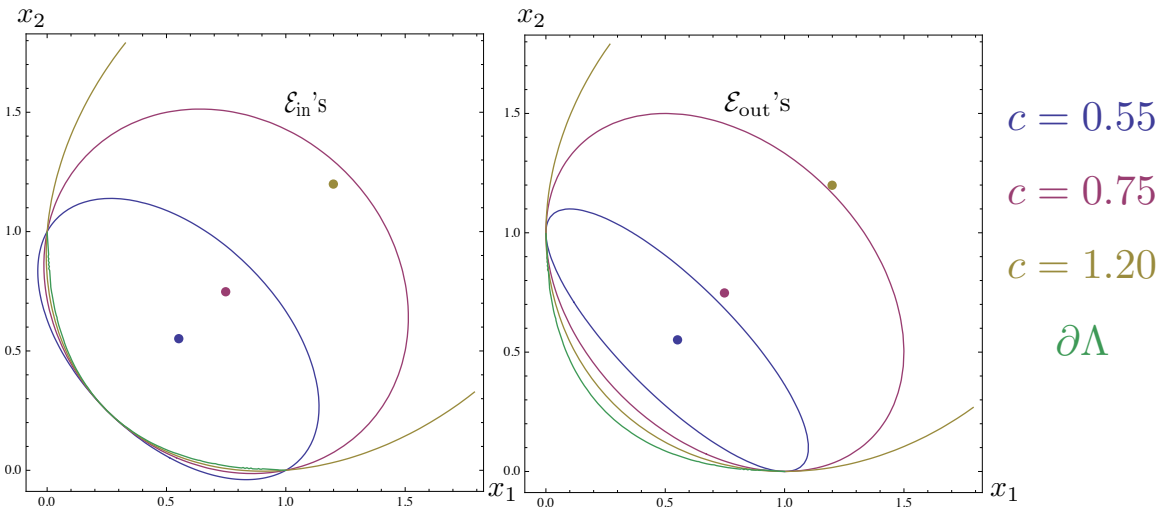

Figure 2.8: Illustration (for $n = 2$) of the fact that tightness of both Λ_{ei} (left) and Λ_{eo} (right) increases in c for $c > 1/n$. For Λ_{ei} the set inclusion relationship is reversed once one goes beyond $\mathcal{S}$, whereas for Λ_{eo} there is a complete set inclusion relationship among this family of ellipsoids. Also shown are $\partial\Lambda$ and the center of each ellipsoid.

sandwiching is asymptotically tight at the all-rates-equal point $\mathbf{m}$ (and by construction always tight at each $\mathbf{e}_i$). Specifically, the ellipsoids $\mathcal{E}_{\text{in}}, \mathcal{E}_{\text{out}}$ viewed in the limit as $c \to \infty$ have axes ratios given by

$$\lim_{c\to\infty} \frac{a_{1,\text{in}}(n,c)}{a_{1,\text{out}}(n,c)} = 1, \quad \lim_{c\to\infty} \frac{a_{2,\text{in}}(n,c)}{a_{2,\text{out}}(n,c)} = \left(2\left(1-\left(1-\frac{1}{n}\right)^{n-1}\right)\right)^{-\frac{1}{2}} \tag{2.120}$$

for each n, where the a_2 axes ratio is a monotone decreasing function of n, starting with value 1 at $n = 2$. Thus when $n = 2$, the $\mathcal{E}_{\text{in}}, \mathcal{E}_{\text{out}}$ are asymptotically equal as $c \to \infty$. It follows that the asymptotic in n and c a_2 axes ratio is

$$\lim_{n\to\infty}\lim_{c\to\infty} \frac{a_{2,\text{in}}(n,c)}{a_{2,\text{out}}(n,c)} = \sqrt{\frac{\text{e}}{2(\text{e}-1)}} \approx 0.8894. \tag{2.121}$$

2.6.3 An alternative proof of the outer bound

This subsection supplies an alternative approach to proving the outer bound in Prop. 17. It originates from this observation: membership testing for a ball is even simpler than that for an ellipsoid in that the Euclidean distance to the center of the ball is the sole indicator of set membership. Also, recall our ellipsoid parameterized by (c, a_1, a_2) is already spherical in the subspace spanned by its

$2^{\text{nd}}, \ldots, n^{\text{th}}$ axes, so the required transformation converting the ellipsoid to a ball is expected to be simple, too. Of course, ultimately we care about bounding Λ rather than membership testing for ellipsoid. The gap is filled in by the following lemma, which uses Prop. 14.

Let $\mathbf{T}_{\text{out}}$ be an affine transformation that transforms $\mathcal{E}_{\text{out}}$ (with center $\mathbf{c} = c\mathbf{1}$ and $c > \frac{1}{n}\left(1 - \frac{1}{n}\right)^{n-1}$) to the unit ball at the origin, i.e., $\mathbf{T}_{\text{out}}\left(\mathcal{E}_{\text{out}}\right) = \mathcal{B}\left(\mathbf{o}, 1\right)$. For any $\mathbf{x}_t \in \partial\Lambda$ denote by $\mathcal{H}_t = \mathcal{H}_t(\mathbf{x}_t)$ the hyperplane tangent to Λ at $\mathbf{x}_t$. In the transformed space, this hyperplane is denoted $\hat{\mathcal{H}}_t = \mathbf{T}_{\text{out}}(\mathcal{H}_t)$.

Lemma 16. *Under the above transformation,*

$$d(\hat{\mathcal{H}}_t, \mathbf{o}) \geq 1 \text{ holds for all } \mathbf{x}_t \in \partial\Lambda \;\Rightarrow\; \Lambda \subseteq \Lambda_{\text{eo}}. \tag{2.122}$$

If $\mathcal{E}_{\text{out}}$ is confined within $\mathbb{R}^n_+$, then we also have a converse, namely

$$\text{If } \mathcal{E}_{\text{out}} \subseteq \mathbb{R}^n_+, \text{ then } \Lambda \subseteq \Lambda_{\text{eo}} \;\Rightarrow\; d(\hat{\mathcal{H}}_t, \mathbf{o}) \geq 1 \text{ holds for all } \mathbf{x}_t \in \partial\Lambda. \tag{2.123}$$

Proof. "forward part (i.e., (2.122))": As affine transformations preserve tangency and set inclusion, it follows that

$$d(\hat{\mathcal{H}}_t, \mathbf{o}) \geq 1 \iff d(\mathcal{H}_t, \overline{\mathcal{E}_{\text{out}}}) > 0 \text{ or } \mathcal{H}_t \text{ is tangent with } \overline{\mathcal{E}_{\text{out}}}. \tag{2.124}$$

The RHS of (2.124) says $\mathcal{H}_t \cap \overline{\mathcal{E}_{\text{out}}} \subseteq \partial\mathcal{E}_{\text{out}}$, since $\mathbf{x}_t \in \mathcal{H}_t$ it follows $\mathbf{x}_t \notin \mathcal{E}_{\text{out}}$. So $d(\hat{\mathcal{H}}_t, \mathbf{o}) \geq 1$ implies $\forall \mathbf{x}_t \in \partial\Lambda,\ \mathbf{x}_t \notin \mathcal{E}_{\text{out}}$, meaning $\partial\Lambda \subseteq \Lambda_{\text{eo}}$. Observe $\partial\Lambda \subseteq \Lambda_{\text{eo}} \iff \partial\Lambda \cap \mathcal{E}_{\text{out}} = \emptyset$, it then follows from Prop. 14 that $\Lambda \subseteq \Lambda_{\text{eo}}$.

"converse part (i.e., (2.123))": $\Lambda \subseteq \Lambda_{\text{eo}}$ means $\mathcal{E}_{\text{out}} \subseteq \Lambda^c$, so $\mathcal{E}_{\text{out}} \subseteq \Lambda^c \cap \mathbb{R}^n_+$, a convex set ([17]). Thus $\mathcal{H}_t$ is a supporting hyperplane for the convex set $\Lambda^c \cap \mathbb{R}^n_+ \supseteq \mathcal{E}_{\text{out}}$, so $d(\mathcal{H}_t, \overline{\mathcal{E}_{\text{out}}}) > 0$ or $\mathcal{H}_t$ is tangent with $\overline{\mathcal{E}_{\text{out}}}$, then the properties of affine transformation ensure $d(\hat{\mathcal{H}}_t, \overline{\mathcal{B}}(\mathbf{o}, 1)) > 0$ or $\hat{\mathcal{H}}_t$ is tangent with $\overline{\mathcal{B}}(\mathbf{o}, 1)$, which can be summarized as $d(\hat{\mathcal{H}}_t, \mathbf{o}) \geq 1$. □

Remark 6. *The intuition behind Lem. 16 is that the intersection between any tangent hyperplane of $\partial\Lambda$ and $\mathbb{R}^n_+$ is included in Λ (proof similar to part of the proof of Prop. 7), hence the intersection*

is disjoint from (or at most tangent with) any $\overline{\mathcal{E}_{\text{out}}}$. *If further* $\mathcal{E}_{\text{out}}$ *is entirely confined within* $\mathbb{R}^n_+$, *then for any* $\mathbf{x}_t \in \partial\Lambda$, *any part of the associated tangent hyperplane* $\mathcal{H}_t(\mathbf{x}_t)$ *beyond* $\mathbb{R}^n_+$ *cannot touch* $\mathcal{E}_{\text{out}}$. *Consequently, the entire tangent hyperplane is disjoint from (or at most tangent with) such* $\overline{\mathcal{E}_{\text{out}}}$.

The following proposition will be used in our alternative proof of the Λ_{eo} bound. For notational convenience, we define following functions, assuming (c, a_1, a_2) is given.

Definition 10.

$$g(\mathbf{p}) \equiv \frac{(n-1)^2}{n}a_1^2 + a_2^2\sum_{i=2}^{n}\left(\frac{1}{\sqrt{n}} - p_i - \frac{1}{1+\sqrt{n}}(1-p_1)\right)^2, \quad f(\mathbf{p}) \equiv \sqrt{g(\mathbf{p})} + \pi(\mathbf{p}) \tag{2.125}$$

$$\tilde{g}(\mathbf{p}) \equiv \frac{(n-1)^2}{n}a_1^2 + \left(-\frac{1}{n} + \sum_{i=1}^{n} p_i^2\right)a_2^2, \qquad \tilde{f}(\mathbf{p}) \equiv \sqrt{\tilde{g}(\mathbf{p})} + \pi(\mathbf{p}) \tag{2.126}$$

It can be verified that

$$g(\mathbf{p}) = \tilde{g}(\mathbf{p}), \quad \text{if} \quad \sum_{i=1}^{n} p_i = 1. \tag{2.127}$$

Proposition 20. *Fix an ellipsoid* $\mathcal{E}(c, a_1, a_2)$. *Define* $\Lambda_{\text{e}} \equiv \mathcal{S} \setminus \mathcal{E}$. *Assume* $c > \frac{1}{n}(1-\frac{1}{n})^{n-1}$. *We have*

$$f(\mathbf{p}) \le c(n-1) \text{ for all } \mathbf{p} \in \partial\mathcal{S} \;\Rightarrow\; \Lambda \subseteq \Lambda_{\text{e}}. \tag{2.128}$$

Conversely,

$$\text{If } \mathcal{E} \subseteq \mathbb{R}^n_+, \text{ then } \Lambda \subseteq \Lambda_{\text{e}} \;\Rightarrow\; f(\mathbf{p}) \le c(n-1) \text{ for all } \mathbf{p} \in \partial\mathcal{S}. \tag{2.129}$$

Proof. Recall all rotation matrices are orthonormal, as one consequence of Prop. 11 the rotation matrix given in (2.93) can be used as the $\mathbf{Q}$ in the decomposition of $\mathbf{R}$ as shown in (2.81). An ellipsoid (in the form (2.81)) can be viewed as the image of a unit ball after some affine transformation:

$$\mathcal{E} = \{\mathbf{x} : \mathbf{x} = \mathbf{c} + \mathbf{A}\mathbf{v} \mid \|\mathbf{v}\| < 1\}, \tag{2.130}$$

where $\mathbf{A}$ is the matrix associated with this affine transformation, and $\mathbf{c}$ is the center of the ellipsoid. It is easy to verify the two forms of an ellipsoid are related by $\mathbf{A} = \mathbf{Q}\mathbf{D}^{\frac{1}{2}}$. Furthermore, the hyperplane

$\left\{\mathbf{x} : \mathbf{n}^\mathsf{T}\mathbf{x} = d\right\}$ corresponds to $\left\{\mathbf{v} : \mathbf{n}_v^\mathsf{T}\mathbf{v} = d_v\right\}$ in the pre-image space $\mathbf{V}$ (where $\mathcal{E}$ becomes a unit ball centered at the origin) and they are related by:

$$\mathbf{n}_v^\mathsf{T} = \mathbf{n}^\mathsf{T}\mathbf{A}, \quad d_v = d - \mathbf{n}^\mathsf{T}\mathbf{c}. \tag{2.131}$$

Recall from[17] that the tangent hyperplane at a point $\mathbf{x}(\mathbf{p})$ on $\partial\Lambda$: $\left\{\mathbf{x} : (\mathbf{1}-\mathbf{p})^\mathsf{T}\mathbf{x} = \pi(\mathbf{p})\right\}$ where $\mathbf{p} \in \partial\mathcal{S}$. Specializing $\mathbf{n} = \mathbf{1} - \mathbf{p}$, $d = \pi(\mathbf{p})$, we can verify:

$$(\mathbf{n}_v)_i = \begin{cases} \frac{1}{\sqrt{n}} a_1 (n-1), & i = 1 \\ a_2 \left(\frac{1}{\sqrt{n}} - p_i - \frac{1}{1+\sqrt{n}}(1-p_1) \right), & i = 2, \ldots, n \end{cases} \tag{2.132}$$

We apply Lem. 16 with $\hat{\mathcal{H}}_t = \left\{\mathbf{v} : \mathbf{n}_v^\mathsf{T}\mathbf{v} = d_v\right\}$, using a formula (see e.g.,[26] §1.4) for computing the absolute distance between a point $\mathbf{y}$ and the hyperplane $\hat{\mathcal{H}}_t$: $|\mathbf{y}^\mathsf{T}\mathbf{n}_v - d_v|/\|\mathbf{n}_v\|$ and observing setting $\mathbf{y} = \mathbf{o}$ and assuming $c > \frac{1}{n}(1-\frac{1}{n})^{n-1}$ allows the sign to be determined. Algebra shows the following equivalent condition:

$$d(\hat{\mathcal{H}}_t, \mathbf{o}) \geq 1 \iff f(\mathbf{p}) \leq c(n-1). \tag{2.133}$$

The proof is now complete. □

Prop. 20 asks us to check the inequality $f(\mathbf{p}) \leq c(n-1)$ holds for all $\mathbf{p} \in \partial\mathcal{S}$. As before, we formulate this as a constrained optimization problem. Unlike before, though, Schur-convexity now makes our job a lot easier. For this we need to state an equivalent condition for verifying Schur-convexity.

Proposition 21 ([27], Ch. 3, Thm. A.4)**.** *Let $\mathcal{D} \subseteq \mathbb{R}$ be an open interval and let a symmetric function $f : \mathcal{D}^n \to \mathbb{R}$ be continuously differentiable. Then f is Schur-convex on $\mathcal{D}^n$ if and only if for all distinct indices $k, l \in [n]$:*

$$(x_k - x_l)\left(\frac{\partial f}{\partial x_k} - \frac{\partial f}{\partial x_l}\right) \geq 0, \ \forall x \in \mathcal{D}^n. \tag{2.134}$$

We now apply Prop. 20, meaning we need to solve

$$\max_{\mathbf{p}\in\partial\mathcal{S}} \ \tilde{f}(\mathbf{p}) = \sqrt{\tilde{g}(\mathbf{p})} + \pi(\mathbf{p}), \tag{2.135}$$

where $g(\mathbf{p})$ ($f(\mathbf{p})$) is replaced by $\tilde{g}(\mathbf{p})$ ($\tilde{f}(\mathbf{p})$) due to (2.127). Note $\tilde{f}(\mathbf{p})$ is symmetric.

Algebra gives:

$$(p_k - p_l)\left(\frac{\partial \tilde{f}}{\partial p_k} - \frac{\partial \tilde{f}}{\partial p_l}\right) = (p_k - p_l)^2 \left(\frac{a_2^2}{\sqrt{\tilde{g}}} - \frac{\pi(\mathbf{p})}{(1-p_k)(1-p_l)}\right). \tag{2.136}$$

In order to show the RHS of (2.136) is non-negative, it suffices to show $\frac{a_2^2}{\sqrt{\tilde{g}}} \geq 1$, for which we specialize with the expressions for $\mathcal{E}_{\text{out}}$ given in (2.114) and get:

$$\left(\frac{a_{2,\text{out}}^2}{\sqrt{\tilde{g}}}\right)^2 = \frac{(n-1)c}{nc - 1 - c + \sum_{i=1}^n p_i^2} \geq \frac{(n-1)c}{nc - 1 - c + \sum_{i=1}^n p_i} = 1. \tag{2.137}$$

Therefore, Prop. 21 tells $\tilde{f}$ in (2.135) specialized to our $\mathcal{E}_{\text{out}}$ is Schur-convex. It then follows from the definition of Schur-convexity and the fact $\mathbf{e}_i$ majorizes every other point in the feasible set $\partial\mathcal{S}$ that the *global* maximum of $\tilde{f}$ is attained at each $\mathbf{e}_i$. Finally, evaluating $\tilde{f}$ at $\mathbf{e}_i$ gives $c(n-1)$, the desired global maximum (according to Prop. 20). This then completes our alternative proof that our proposed ellipsoid outer bound is valid.

Remark 7. *The Schur-convexity approach would not be directly applicable to the proposed ellipsoid inner bound in Prop. 18. Because even if we could have a parallel result to Prop. 20 for inner bounding ellipsoids, the fact that our inner bounding ellipsoid in* (2.115) *is tight at* $\mathbf{m}$ *as well as at* $\mathbf{e}_i$ *precludes Schur-convexity, since* $\mathbf{m}$ *is strictly majorized by every other point and* $\mathbf{e}_i$ *majorizes every other point.*

2.7 Generalized convexity properties

Given an arrival rate vector $\mathbf{x} = (x_1, \ldots, x_n)$, define the set of stabilizing controls $\mathcal{P}(\mathbf{x})$ assuming the worst-case service rate:

$$\mathcal{P}(\mathbf{x}) \equiv \left\{ \mathbf{p} \in [0,1]^n : x_i \leq p_i \prod_{j \neq i} (1 - p_j),\ i \in [n] \right\}. \tag{2.138}$$

This set is related to Λ in that $\mathcal{P}(\mathbf{x}) = \emptyset$ iff $\mathbf{x} \notin \Lambda$. Whereas Λ is an important inner bound on the Aloha stability region Λ_{A}, the set $\mathcal{P}(\mathbf{x})$ can be viewed as the set of control options given a desired arrival rate vector $\mathbf{x}$. To proceed, we now view $\mathbf{x}$ as parameters (instead of $\mathbf{p}$, as was done previously) and define an "excess rate" function for each $i \in [n]$:

$$f_i(\mathbf{p}) \equiv x_i - p_i \prod_{j \neq i} (1 - p_j),\ i \in [n]. \tag{2.139}$$

The excess rate functions are closely related to the set of stabilizing controls. To see this, define the α-sublevel sets of each excess rate function

$$S_{i,\alpha} \equiv \{ \mathbf{p} \in [0,1]^n : f_i(\mathbf{p}) \leq \alpha \},\ \alpha \in \mathbb{R}. \tag{2.140}$$

Denote by $S_\alpha = \cap_{i=1}^n S_{i,\alpha}$; then $\mathcal{P}(\mathbf{x}) = S_0 = \cap_{i=1}^n S_{i,0}$. In words, the set of stabilizing controls associated with Λ is the intersection of 0-sublevel sets of these n excess rate functions.

For simplicity and sometimes, well-definedness, throughout this section we will work with the open convex domain $(0,1)^n$. The following propositions show these excess rate functions are not convex (Prop. 22), but are quasiconvex (Prop. 23), pseudoconvex (Prop. 24), and invex (Prop. 25).

Proposition 22. *For all $i \in [n]$, the excess rate function $f_i(\mathbf{p})$ in (2.139) is not convex on $(0,1)^n$.*

Proof. In fact we can work with an additional constraint $\sum_{i=1}^n p_i = 1$, i.e., with $\partial\mathcal{S}$ being the domain. W.l.o.g. let us show this for $f_1(\mathbf{p}) = x_1 - p_1 \prod_{j \neq 1}(1 - p_j)$. Let $\epsilon \in (0,1)$ be determined later. Denote $\mathbf{p}_{1,\epsilon} = (1 - \epsilon, \epsilon/(n-1), \ldots, \epsilon/(n-1))$, $\mathbf{p}_2 = (1/n)\mathbf{1}$. Form the convex combination

$\mathbf{p}_{\theta,\epsilon} = \theta\mathbf{p}_{1,\epsilon} + (1-\theta)\mathbf{p}_2$ for $\theta \in (0,1)$. We shall show for all n the existence of ϵ and θ in order for the following inequality to hold:

$$f_1(\mathbf{p}_{\theta,\epsilon}) > \theta f_1(\mathbf{p}_{1,\epsilon}) + (1-\theta) f_1(\mathbf{p}_2). \tag{2.141}$$

After substituting definitions and $\theta = 1/2$, the above inequality becomes:

$$(1-\epsilon+\frac{1}{n})\left(1-\frac{n(\epsilon+1)-1}{2n(n-1)}\right)^{n-1} < (1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)^{n-1} + \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}. \tag{2.142}$$

Manipulation of the above equation gives the equivalent form

$$\sqrt[n-1]{\frac{1-\epsilon+\frac{1}{n}}{(1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)^{n-1} + \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}}} < \frac{1}{1-\frac{n(\epsilon+1)-1}{2n(n-1)}}. \tag{2.143}$$

Applying the AM-GM inequality to the LHS:

$$\sqrt[n-1]{\frac{1-\epsilon+\frac{1}{n}}{(1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)^{n-1} + \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}} \cdot \underbrace{1\cdots 1}_{\#:n-2}} \leq \frac{\frac{1-\epsilon+\frac{1}{n}}{(1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)^{n-1} + \frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}} + n - 2}{n-1}. \tag{2.144}$$

It is easily verified that the sequence $\left(1-\frac{1}{n}\right)^{n-1}$ monotonically decreases to $1/\mathrm{e}$, and that $\left(1-\frac{\epsilon}{n-1}\right)^{n-2}$ monotonically decreases to $1/\mathrm{e}^{\epsilon}$. Because of this, it suffices to show

$$\frac{\frac{1-\epsilon+\frac{1}{n}}{(1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)\frac{1}{\mathrm{e}^{\epsilon}} + \frac{1}{n}\frac{1}{\mathrm{e}}} + n - 2}{n-1} < \frac{1}{1-\frac{n(\epsilon+1)-1}{2n(n-1)}}, \tag{2.145}$$

which after rearrangement becomes

$$\frac{1-\epsilon+\frac{1}{n}}{(1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)\frac{1}{\mathrm{e}^{\epsilon}} + \frac{1}{n}\frac{1}{\mathrm{e}}} - \frac{3n-2}{2n-1} < 0. \tag{2.146}$$

Denote $h_1(n,\epsilon) = \frac{1-\epsilon+\frac{1}{n}}{(1-\epsilon)\left(1-\frac{\epsilon}{n-1}\right)\frac{1}{\mathrm{e}^{\epsilon}} + \frac{1}{n}\frac{1}{\mathrm{e}}}$ and $h_2(n) = \frac{3n-2}{2n-1}$. One can verify that, given $\epsilon \in (0,1)$, $h_1(n,\epsilon)$ and $h_2(n)$ are monotonically decreasing and increasing in n ($n \geq 2$), respectively. Thus

it suffices to show (2.146) holds when $n = 2$. Observe that given n, the LHS of (2.146) i.e., $h_1(n,\epsilon) - h_2(n)$ is a continuous function of ϵ for $\epsilon \in (0,1)$, since $h_1(2,0) - h_2(2) < 0$, $h_1(2,1) - h_2(2) > 0$, there exist various choices of ϵ so that (2.146) holds for $n = 2$, and hence for all $n \geq 2$ (with the same choice of ϵ). □

Recall a function is called quasiconvex (or unimodal) if its domain and all its sublevel sets are convex[23].

Proposition 23. *The excess rate function is quasiconvex on* $(0,1)^n$ *for all* $i \in [n]$.

Proof. Our approach is to show the convexity of the sublevel sets $S_{i,\alpha}$ for which we discuss two cases: *i)* $\alpha \geq x_i$, and *ii)* $\alpha < x_i$. Consider case *i)*. If $\alpha \geq x_i$ then $f_i(\mathbf{p}) \leq \alpha$ always holds, and therefore $S_{i,\alpha} = \mathrm{dom} f = (0,1)^n$, which is convex. It remains to consider case *ii)*, with $\alpha < x_i$. Construct $g_i(\mathbf{p}_{\setminus i}) \equiv \frac{x_i - \alpha}{\prod_{j \neq i}(1-p_j)}$, where $\mathbf{p}_{\setminus i}$ is formed from the vector $\mathbf{p}$ by dropping component i. Observe

$$S_{i,\alpha} = \{\mathbf{p} \in (0,1)^n : f_i(\mathbf{p}) \leq \alpha\} = \{(\mathbf{p}_{\setminus i}, p_i) \in (0,1)^n : g_i(\mathbf{p}_{\setminus i}) \leq p_i\}. \tag{2.147}$$

Thus $S_{i,\alpha}$ can be interpreted as the *epigraph* of the function $g_i(\mathbf{p}_{\setminus i})$. Since a function is convex iff its epigraph is a convex set, we then need to show this function $g_i(\mathbf{p}_{\setminus i})$ is convex. Towards this, we take the logarithm and write:

$$\log g_i(\mathbf{p}_{\setminus i}) = \log(x_i - \alpha) - \sum_{j \neq i} \log(1 - p_j) = \log(x_i - \alpha) + \sum_{j \neq i} -\log\left(1 - (\mathbf{e}_j)_{\setminus i}^{\mathsf{T}} \mathbf{p}_{\setminus i}\right), \tag{2.148}$$

where $(\mathbf{e}_j)_{\setminus i}$ is the $(n-1)$-vector by peeling off the i^{th} component of $\mathbf{e}_j$. This shows the RHS of the above equation is convex by recognizing the convexity of $-\log(\cdot)$ and certain function compositions that preserve convexity. Finally since the function $g_i(\mathbf{p}_{\setminus i})$ is log-convex, this means $g_i(\mathbf{p}_{\setminus i})$ is itself convex, which means its epigraph, or equivalently the sublevel set $S_{i,\alpha}$, is convex. □

Corollary 3. *Recall* $\mathcal{P}(\mathbf{x}) = S_0 = \cap_{i=1}^n S_{i,0}$. *It follows that the set of stabilizing controls* $\mathcal{P}(\mathbf{x})$ *associated with* $\mathbf{x} \in \Lambda$ *is convex, as convexity is preserved under set intersection.*

Proposition 24. *The excess rate function is pseudoconvex on* $(0,1)^n$ *for all* $i \in [n]$.

Proof. We apply Cor. 10.1 of Diewert et al.[28] (note it is stated for pseudoconcavity), or Cambini and Martein[29] (Thm. 3.4.6 and 3.4.7). Essentially, we shall show that given $i \in [n]$ and $\mathbf{p} \in \mathrm{dom} f = (0,1)^n$, then for all $\mathbf{q}$ such that $\|\mathbf{q}\| = 1$, $\mathbf{q}^\mathsf{T} \nabla f_i(\mathbf{p}) = 0$ implies $\mathbf{q}^\mathsf{T} \nabla^2 f_i(\mathbf{p}) \mathbf{q} > 0$.

The gradient of $f_i(\mathbf{p})$ is:

$$\frac{\partial}{\partial p_k} f_i(p) = \begin{cases} \frac{f_i(p)-x_i}{p_i}, & k = i \\ -\frac{f_i(p)-x_i}{1-p_k}, & k \neq i \end{cases} = \begin{cases} -\prod_{j \neq i}(1-p_j), & k = i \\ p_i \prod_{j \neq i,k}(1-p_j), & k \neq i \end{cases}, \; i \in [n]. \tag{2.149}$$

The Hessian of $f_i(\mathbf{p})$ is:

$$\frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) = \begin{cases} 0, & k = l \\ \frac{f_i(\mathbf{p})-x_i}{(1-p_k)(1-p_l)}, & (i,k,l) \text{ distinct} \\ -\frac{f_i(\mathbf{p})-x_i}{p_i(1-p_l)}, & k = i \neq l \\ -\frac{f_i(\mathbf{p})-x_i}{p_i(1-p_k)}, & l = i \neq k \end{cases} = \begin{cases} 0, & k = l \\ -p_i \prod_{j \neq k,l,i}(1-p_j), & (i,k,l) \text{ distinct} \\ \prod_{j \neq l,i}(1-p_j), & k = i \neq l \\ \prod_{j \neq k,i}(1-p_j), & l = i \neq k \end{cases}, \; i \in [n]. \tag{2.150}$$

Given $\mathbf{q}$, $\mathbf{q}^\mathsf{T} \nabla f_i(\mathbf{p}) = 0$ means

$$q_i \frac{f_i(\mathbf{p}) - x_i}{p_i} + \sum_{k \neq i} q_k \left(-\frac{f_i(\mathbf{p}) - x_i}{1 - p_k} \right) = 0, \tag{2.151}$$

or equivalently:

$$q_i \left(-\prod_{j \neq i}(1-p_j) \right) + \sum_{k \neq i} q_k \left(p_i \prod_{j \neq k,i} (1-p_j) \right) = 0, \tag{2.152}$$

which gives (since $\mathbf{p} \in (0,1)^n$):

$$\sum_{k \neq i} \frac{q_k}{1-p_k} = \frac{q_i}{p_i}. \tag{2.153}$$

To verify $\mathbf{q}^\mathsf{T} \nabla^2 f_i(\mathbf{p}) \mathbf{q} > 0$, i.e., $\sum_{k,l} q_k q_l \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) > 0$, we break this sum into four parts. For any given $i \in [n]$, the set $\mathcal{I}_i \equiv \{(k,l) : k,l \in [n]\}$ may be partitioned into four parts:

$$\mathcal{I}_{i,1} = \{(k,k) : k \in [n]\},\ \mathcal{I}_{i,2} = \{(k,l) : k,l,i \text{ distinct}\},\ \mathcal{I}_{i,3} = \{(i,l) : l \in [n] \backslash \{i\}\},\ \mathcal{I}_{i,4} = \{(k,i) : k \in [n] \backslash \{i\}\}. \tag{2.154}$$

Then

$$\sum_{k,l} q_k q_l \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) = \sum_{(k,l)\in\mathcal{I}_{i,1}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) + \sum_{(k,l)\in\mathcal{I}_{i,2}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) + \sum_{(k,l)\in\mathcal{I}_{i,3}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) + \sum_{(k,l)\in\mathcal{I}_{i,4}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}).$$

By symmetry, the third and fourth sums are equal to each other. The three sums equal, respectively:

$$\sum_{(k,l)\in\mathcal{I}_{i,1}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) = 0 \tag{2.155}$$

$$\sum_{(k,l)\in\mathcal{I}_{i,2}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) = \sum_{(k,l)\in\mathcal{I}_{i,2}} q_k q_l \left(\frac{f_i(\mathbf{p}) - x_i}{(1-p_k)(1-p_l)} \right) \tag{2.156}$$

$$\sum_{(k,l)\in\mathcal{I}_{i,3}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) = \sum_{(k,l)\in\mathcal{I}_{i,3}} q_k q_l \left(-\frac{f_i(\mathbf{p}) - x_i}{p_i(1-p_l)} \right) = \sum_{l\neq i} q_i q_l \left(-\frac{f_i(\mathbf{p}) - x_i}{p_i(1-p_l)} \right) \tag{2.157}$$

The third sum can be further simplified as

$$\begin{aligned} \sum_{(k,l)\in\mathcal{I}_{i,3}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) &= q_i \frac{1}{p_i} \sum_{l\neq i} q_l \left(-\frac{f_i(\mathbf{p}) - x_i}{1-p_l} \right) \\ &\overset{(a)}{=} q_i \frac{1}{p_i} \left(-q_i \frac{f_i(\mathbf{p}) - x_i}{p_i} \right) \\ &= -\left(\frac{q_i}{p_i} \right)^2 (f_i(\mathbf{p}) - x_i) \\ &= \left(\frac{q_i}{p_i} \right)^2 p_i \prod_{j\neq i} (1-p_j), \end{aligned} \tag{2.158}$$

where (a) is due to (2.151). Next, the second sum can be written as:

$$\sum_{(k,l)\in\mathcal{I}_{i,2}} \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) = \sum_{(k,l)\in\mathcal{I}_{i,2}} q_k q_l \left(-p_i \prod_{j\neq k,l,i} (1-p_j) \right) = -\sum_{(k,l)\in\mathcal{I}_{i,2}} q_k q_l p_i \frac{\prod_{j\neq i}(1-p_j)}{(1-p_k)(1-p_l)}. \tag{2.159}$$

Combining the above results, we have

$$\begin{aligned} \sum_{k,l} q_k q_l \frac{\partial^2}{\partial p_k \partial p_l} f_i(\mathbf{p}) &= -\sum_{(k,l)\in\mathcal{I}_{i,2}} q_k q_l p_i \frac{\prod_{j\neq i}(1-p_j)}{(1-p_k)(1-p_l)} + 2\left(\frac{q_i}{p_i} \right)^2 p_i \prod_{j\neq i}(1-p_j) \\ &= p_i \prod_{j\neq i}(1-p_j) \left(-\sum_{(k,l)\in\mathcal{I}_{i,2}} \frac{q_k q_l}{(1-p_k)(1-p_l)} + 2\left(\frac{q_i}{p_i} \right)^2 \right). \end{aligned} \tag{2.160}$$

Then, the sign of the overall sum is determined by the sign of the term in parentheses above. This term can be written as

$$\begin{aligned} -\sum_{(k,l)\in\mathcal{I}_{i,2}} \frac{q_k q_l}{(1-p_k)(1-p_l)} + 2\left(\frac{q_i}{p_i}\right)^2 &= -\left(\sum_{k\neq i}\frac{q_k}{1-p_k}\sum_{l\neq i}\frac{q_l}{1-p_l} - \sum_{k=l\neq i}\frac{q_k}{1-p_k}\frac{q_l}{1-p_l}\right) + 2\left(\frac{q_i}{p_i}\right)^2 \\ &\overset{(b)}{=} -\left(\left(\frac{q_i}{p_i}\right)^2 - \sum_{k\neq i}\left(\frac{q_k}{1-p_k}\right)^2\right) + 2\left(\frac{q_i}{p_i}\right)^2 \\ &= \left(\frac{q_i}{p_i}\right)^2 + \sum_{k\neq i}\left(\frac{q_k}{1-p_k}\right)^2, \end{aligned} \tag{2.161}$$

where (b) is due to (2.153), and the last expression is clearly positive. This establishes the pseudo-convexity of the excess rate function $f_i(\mathbf{p})$ on $\mathbf{p} \in (0,1)^n$. □

Proposition 25. *The excess rate function is invex on* $(0,1)^n$ *for all* $i \in [n]$.

Proof. For differentiable f with open convex domain, f is invex if and only if every stationary point is a global minimizer (see e.g., Thm. 4.9.1 of Cambini and Martein[29]). Next, Thm. 2.27 of Mishra and Giorgi[30] says if f is differentiable and quasiconvex with open convex domain, then f is pseudoconvex if and only if every stationary point is a global minimizer. These two results mean that under the assumption of quasiconvexity, invexity and pseudoconvexity coincide. Thus the invexity of our excess rate functions follows from Props. 23 and 24. □

2.8 Appendix

2.8.1 Proof of Prop. 3

We first show a supporting lemma and its corollary. We define a univariate function $g(t_\pi) = g(t_\pi, \mathbf{p})$, as

$$g(t_\pi, \mathbf{p}) \equiv \prod_{i=1}^{n}(1+p_i t_\pi) - (1+t_\pi),\ \ \mathbf{p} \in [0,1]^n. \tag{2.162}$$

Lemma 17. *For all* $n \geq 2$, $g(t_\pi, \mathbf{p})$ *can only have one or two real roots on* $(-1, \infty)$. *More specifically:*

- $t_\pi = 0$ *is always a root, and is the unique root iff* $\sum_{i=1}^n p_i = 1$, *i.e.,* $\mathbf{p}$ *is a probability vector;*
- *besides* $t_\pi = 0$, *the other root is on* $(0, \infty)$ *iff* $\sum_{i=1}^n p_i < 1$;

- *besides $t_\pi = 0$, the other root is on $(-1, 0)$ iff $\sum_{i=1}^n p_i > 1$.*

Proof. Applying the chain rule of differentiation, we have:

$$g'(t_\pi) = \sum_{i=1}^n p_i \prod_{j\neq i}(1 + p_j t_\pi) - 1, \quad g''(t_\pi) = \sum_{i=1}^n p_i \sum_{j\neq i} p_j \prod_{k\neq j,i} (1 + p_k t_\pi). \tag{2.163}$$

Here are some simple yet important observations: $g(0) = 0$, $g(-1) = \pi(\mathbf{p}) > 0$, $g'(0) = \sum_{i=1}^n p_i - 1$, and $g'(-1) = \sum_{i=1}^n p_i \prod_{j\neq i}(1 - p_j) - 1 \leq 0$. The last inequality is justified since we can construct a vector $\mathbf{x}(\mathbf{p})$ according to (2.9) in Def. 2, and then apply the fact $\Lambda_{\text{eq}} = \Lambda \subseteq \mathcal{S}$. Furthermore, $g''(0) \geq 0$ and in fact $g''(t_\pi) \geq 0$ for all $t_\pi > -1$ which means $g'(t_\pi)$ is monotone increasing on $(-1, \infty)$. In the following we first show the forward part, i.e., how to go from the condition on $\sum_{i=1}^n p_i$ to the properties of the roots.

- case 1: $\sum_{i=1}^n p_i = 1$. In this case, since $g'(0) = 0$ and $g''(0) \geq 0$, the stationary point 0 is a local minimizer. As $g(-1) = \pi(\mathbf{p})$, $g'(-1) \leq 0$, $g'(t_\pi)$ is monotone increasing on $(-1, \infty)$, so what happens on $(-1, \infty)$ is: $g(t_\pi)$ is monotone decreasing from $\pi(\mathbf{p})$ ($t_\pi = -1$) to 0 ($t_\pi = 0$), and then monotone increasing from 0 to ∞ (as $t_\pi \to \infty$). Thus the only root on $(-1, \infty)$ is 0.

- case 2: $\sum_{i=1}^n p_i < 1$. In this case, $g'(0) < 0$. Thus $g(t_\pi)$ is decreasing from $\pi(\mathbf{p})$ ($t_\pi = -1$) to 0 ($t_\pi = 0$), then keeps decreasing until some stationary point $t^*_{\pi,2} > 0$ such that $g'(t^*_{\pi,2}) = 0$, after which $g(t_\pi)$ keeps increasing as $t_\pi \to \infty$. Note $g(t^*_{\pi,2}) < 0$, so the only other root $\tilde{t}_{\pi,2}$ is on $(0, \infty)$.

- case 3: $\sum_{i=1}^n p_i > 1$. In this case, $g'(0) > 0$. Thus $g(t_\pi)$ is decreasing from $\pi(\mathbf{p})$ ($t_\pi = -1$) to 0 (at some $\tilde{t}_{\pi,3}$), and keeps decreasing until at some stationary point $t^*_{\pi,3}$ such that $g'(t^*_{\pi,3}) = 0$, after which $g(t_\pi)$ keeps increasing as $t_\pi \to \infty$. Since $g'(t_\pi) > 0$ for all $t_\pi > t^*_{\pi,3}$ and recall $g'(t_\pi)$ itself is monotone increasing, so $t^*_{\pi,3} \in (-1, 0)$ which further implies the other root i.e., $\tilde{t}_{\pi,3}$ has to be on $(-1, 0)$.

The converses are clear (proof by contradiction) and are omitted. □

Corollary 4. *Fix $n \geq 2$ and $\mathbf{x} \in \mathbb{R}^n_+$. i) The total number of positive roots of $f(\delta, \mathbf{x})$ is at most two. Furthermore, ii) if there exists a compatible stabilizing control $\mathbf{p}$ with $\mathbf{x}$ in the sense of Λ_{eq}, namely the $(\mathbf{x}, \mathbf{p})$ pair satisfies (2.9) in Def. 2, then $f(\delta, \mathbf{x})$ has a positive root δ_t iff $g(t_\pi, \mathbf{p})$ has a root $t_{\pi_\mathbf{p}} \in (-1, \infty)$, and the roots of f, g can be related by $\delta_t = \frac{1}{\pi(\mathbf{p})}(1 + t_{\pi_\mathbf{p}})$. More specifically $f(\delta, \mathbf{x})$ has either one or two positive roots:*

- *$\delta_t = \frac{1}{\pi(\mathbf{p})}$ is always a positive root, and is the unique root iff $\sum_{i=1}^n p_i = 1$, i.e., $\mathbf{p}$ is a probability vector;*
- *besides $\delta_t = \frac{1}{\pi(\mathbf{p})}$, $f(\delta, \mathbf{x})$ also has a larger positive root iff $\sum_{i=1}^n p_i < 1$;*
- *besides $\delta_t = \frac{1}{\pi(\mathbf{p})}$, $f(\delta, \mathbf{x})$ also has a smaller positive root iff $\sum_{i=1}^n p_i > 1$.*

Proof. We first show *ii)*. Substituting δ_t, $\pi(\mathbf{p})$ and $x_i(\mathbf{p}) = \frac{p_i}{1-p_i}\pi(\mathbf{p})$ into the definition of $f(\delta, \mathbf{x})$, we have

$$\begin{aligned} f(\delta_t, \mathbf{x}) &= \prod_{i=1}^n \left(1 + \frac{p_i}{1-p_i}\pi(\mathbf{p})\frac{1}{\pi(\mathbf{p})}(1 + t_{\pi_\mathbf{p}})\right) - \frac{1}{\pi(\mathbf{p})}(1 + t_{\pi_\mathbf{p}}) \\ &= \frac{1}{\pi(\mathbf{p})}\left(\prod_{i=1}^n (1 + p_i t_{\pi_\mathbf{p}}) - (1 + t_{\pi_\mathbf{p}})\right) = \frac{1}{\pi(\mathbf{p})} g(t_{\pi_\mathbf{p}}, \mathbf{p}). \end{aligned} \tag{2.164}$$

Therefore, given $\mathbf{x}$ and its compatible $\mathbf{p}$ in the sense of Λ_{eq}, δ_t is a root of $f(\delta, \mathbf{x})$ iff $t_{\pi_\mathbf{p}}$ is a root of $g(t_\pi, \mathbf{p})$. Since δ_t is positive iff $t_{\pi_\mathbf{p}} \in (-1, \infty)$, the statement for the three cases then follows from Lem. 17.

We next show *i)*. Recall from the proof of Prop. 1 that if $f(\delta, \mathbf{x})$ ever has a positive root $\tilde{\delta}$, then we can construct $\mathbf{p} = \mathbf{p}(\tilde{\delta}, \mathbf{x})$ according to (2.7) in Def. 2. Since this $\mathbf{p}$ is compatible with $\mathbf{x}$ in the sense of Λ_{eq} meaning $(\mathbf{x}, \mathbf{p}(\tilde{\delta}, \mathbf{x}))$ satisfies (2.9) in Def. 2, the assertion that $f(\delta, \mathbf{x})$ can have at most two positive roots follows from *ii)* just proved. □

We now provide the proof of Prop. 3.

Proof of part 1

"⇒":

Given $\mathbf{x} \in \partial\Lambda$, from the (preliminary) root testing Prop. 1 we know there exists a positive root δ_u of $f(\delta, \mathbf{x})$. Form a vector of probabilities $\mathbf{p}_u = \mathbf{p}_u(\delta_u, \mathbf{x})$ according to (2.7) in Def. 2; it can be verified $(\mathbf{x}, \mathbf{p}_u)$ satisfies (2.9) in Def. 2 namely $\mathbf{p}_u$ is compatible with $\mathbf{x}$ in the sense of Λ_{eq}. Massey and Mathys[16] established the one-to-one correspondence (i.e., bijection) between $\partial\mathcal{S}$ and $\partial\Lambda$, so $\mathbf{p}_u \in \partial\mathcal{S}$ and is the only (critical) stabilizing control for $\mathbf{x}$. It can also be verified that $\frac{1}{\pi(\mathbf{p}_u)}$ is a positive root of $f(\delta, \mathbf{x})$. That the root is unique follows from case 1 in Cor. 4.

"$\Leftarrow$":

If $f(\delta, \mathbf{x})$ has a unique positive root denoted δ_u, again form a vector of probabilities $\mathbf{p}_u = \mathbf{p}_u(\delta_u, \mathbf{x})$ according to (2.7), which is compatible with $\mathbf{x}$ in the sense of Λ_{eq}. It follows from Cor. 4 that $g(t_\pi, \mathbf{p}_u)$ has a unique root on $(-1, \infty)$ denoted $t_{\pi_{\mathbf{p}_u}} = 0$, and is related to δ_u by $\delta_u = \frac{1}{\pi(\mathbf{p}_u)}(1 + t_{\pi_{\mathbf{p}_u}}) = \frac{1}{\pi(\mathbf{p}_u)}$. Furthermore $\mathbf{p}_u$ is a probability vector (Lem. 17), which implies $\mathbf{x} \in \partial\Lambda$[16] .

Proof of part 2

Given $\mathbf{x} \in \Lambda \setminus \partial\Lambda$, there must exist (Prop. 1) a positive root δ of $f(\delta_t, \mathbf{x})$. We use this δ to construct a vector of probabilities $\mathbf{p} = \mathbf{p}(\delta, \mathbf{x})$ as in (2.7). Recall then the root can be expressed as $\delta = \frac{1}{\pi(\mathbf{p})}$, and furthermore $(\mathbf{x}, \mathbf{p})$ satisfies (2.9) in Def. 2. Now we use $\mathbf{p}$ to form $g(t_\pi, \mathbf{p})$ and solve for roots on $(-1, \infty)$. From part 1 of this proposition and Cor. 4 we know there must exist a root (denoted t'_π) other than 0 on $(-1, \infty)$. Define $\delta' \equiv \frac{1}{\pi(\mathbf{p})}(1 + t'_\pi)$, with which we further define $\mathbf{p}' = \mathbf{p}'(\delta', \mathbf{x})$, $\mathbf{x}' = \mathbf{x}'(\mathbf{p}')$ as in Def. 2. First observe δ' is a positive root of $f(\delta_t, \mathbf{x})$ due to Cor. 4 (since t'_π solves $g(t_\pi, \mathbf{p}) = 0$). Second we claim $\mathbf{x}' = \mathbf{x}$, because

$$\pi(\mathbf{p}') = \prod_{i=1}^{n}(1 - p'_i) = \prod_{i=1}^{n}\left(1 - \frac{\delta' x_i}{1 + \delta' x_i}\right) = \frac{1}{\prod_{i=1}^{n}(1 + \delta' x_i)} = \frac{1}{\delta'}, \tag{2.165}$$

where the last equality follows again from Cor. 4. It can then be verified that $x'_i = \frac{p'_i}{1-p'_i}\pi(\mathbf{p}') = x_i$ for all i.

Note $\delta' \neq \delta$ (as $t'_\pi \neq 0$), and as such we can repeat the above procedure starting from this different positive root δ'. More specifically, define $\mathbf{p}'_r = \mathbf{p}'_r(\delta', \mathbf{x}')$ as in (2.7). Then the root satisfies $\delta' = \frac{1}{\pi(\mathbf{p}'_r)}$ and furthermore $(\mathbf{x}', \mathbf{p}'_r)$ satisfies (2.9) in Def. 2. We now use $\mathbf{p}'_r$ to form $g(t_\pi, \mathbf{p}'_r)$ and solve for the root other than 0 on $(-1, \infty)$. We claim this root has to be such that it allows us to

reconstruct $\mathbf{p}$. To see this, denote this root as $t''_\pi \in (-1,0)\cup(0,\infty)$. Define $\delta'' \equiv \frac{1}{\pi(\mathbf{p}'_r)}(1+t''_\pi)$, with which we further define $\mathbf{p}''_r = \mathbf{p}''_r(\delta'', \mathbf{x}')$, $\mathbf{x}''_r = \mathbf{x}''_r(\mathbf{p}''_r)$ as in Def. 2. According to Cor. 4, $f(\delta'', \mathbf{x}') = \frac{1}{\pi(\mathbf{p}'_r)} g(t''_\pi, \mathbf{p}'_r)$, then δ'' is a positive root of $f(\delta_t, \mathbf{x}') = 0$. Now since $\mathbf{x}' = \mathbf{x}$ (so $f(\delta_t, \mathbf{x}')$ and $f(\delta_t, \mathbf{x})$ are the same) and furthermore f has exactly two positive roots (since Cor. 4 says f can have at most two), it then has to hold that $\delta'' = \delta$ (as $t''_\pi \neq 0$ implies $\delta'' \neq \delta'$), which further gives $\mathbf{p}''_r = \mathbf{p}$, $\pi(\mathbf{p}''_r) = \pi(\mathbf{p})$ and $\mathbf{x}''_r = \mathbf{x}$. This proves the claim. Effectively this says the above procedure (as illustrated in (2.166) below) can be reversed.

Furthermore, if $\mathbf{p}$ is such that $\sum_{i=1}^n p_i < 1$, then $\mathbf{p}'$ is such that $\sum_{i=1}^n p'_i > 1$, and vice versa. To show this, assume wlog $\mathbf{p}$ is such that $\sum_{i=1}^n p_i < 1$. Now for $\delta = \frac{1}{\pi(\mathbf{p})}$, $\delta' = \frac{1}{\pi(\mathbf{p})}(1+t'_\pi)$, Lem. 17 says if $\sum_{i=1}^n p_i < 1$ then $t'_\pi > 0$ and hence $\delta' > \delta$. From (2.165) δ' can also be expressed as $\frac{1}{\pi(\mathbf{p}')}$ (so $\pi(\mathbf{p}') < \pi(\mathbf{p})$). Now let us repeat the above procedure as illustrated in diagram (2.166) starting from δ' (rather than δ), we will then get $\delta'' = \frac{1}{\pi(\mathbf{p}')}(1+t''_\pi)$ where t''_π is a root on $(-1,0)\cup(0,\infty)$ solved from $g(t_\pi, \mathbf{p}') = 0$. Since it has to hold that $\delta'' = \delta = \frac{1}{\pi(\mathbf{p})}$, it then follows $t''_\pi < 0$ which in turn implies (Lem. 17) $\sum_{i=1}^n p'_i > 1$. This shows, as a result of the above procedure one of the two critical stabilizing controls is in $\mathcal{S} \setminus \partial\mathcal{S}$, the other in $[0,1]^n \setminus \mathcal{S}$.

Finally we show there can not be more than two critical stabilizing controls for a given $\mathbf{x}$. We prove by contradiction. Assume wlog there are two critical stabilizing controls $\mathbf{p}_s$ and $\tilde{\mathbf{p}}_s$ both in $\mathcal{S}$ (for the case of $\mathcal{S}^c$ the proof is similar), since both $\mathbf{p}_s$ and $\tilde{\mathbf{p}}_s$ are critical stabilizing controls for the same $\mathbf{x}$ it has to hold that $\pi(\mathbf{p}_s) \neq \pi(\tilde{\mathbf{p}}_s)$ (as otherwise $\mathbf{p}_s$ would not be distinct from $\tilde{\mathbf{p}}_s$). As we have just shown above for both $\mathbf{p}_s$ and $\tilde{\mathbf{p}}_s$ there is a corresponding vector of probabilities in $[0,1]^n \setminus \mathcal{S}$ that critically stabilizes $\mathbf{x}$, denoted as $\mathbf{p}_{s^c}$ and $\tilde{\mathbf{p}}_{s^c}$ respectively. From the previous parts of the proof we see it has to hold that $\frac{1}{\pi(\mathbf{p}_s)} \neq \frac{1}{\pi(\mathbf{p}_{s^c})}$ and $\frac{1}{\pi(\tilde{\mathbf{p}}_s)} \neq \frac{1}{\pi(\tilde{\mathbf{p}}_{s^c})}$. This means $f(\delta_t, \mathbf{x})$ already has more than two positive roots, which is impossible according to Cor. 4 (recall with the exceptions of $\mathbf{e}_i$'s any critical stabilizing control $\mathbf{p}$ gives a positive root of f as $\frac{1}{\pi(\mathbf{p})}$).

$$\underset{\mathbf{x}\in\Lambda\setminus\partial\Lambda}{\text{given}} \xrightarrow{\substack{\text{solve}\\ f(\delta_t,\mathbf{x})=0}} \underset{\delta>0}{\text{root}} \longrightarrow \underset{\text{as in Def. 2}}{\text{set } \mathbf{p}=\mathbf{p}(\delta,\mathbf{x})} \longrightarrow \underset{\text{as in Def. 2}}{\delta=\tfrac{1}{\pi(\mathbf{p})},\ \mathbf{x}=\mathbf{x}(\mathbf{p})} \xrightarrow{\substack{\text{solve}\\ g(t_\pi,\mathbf{p})=0}} \underset{t'_\pi\in(-1,0)\cup(0,\infty)}{\text{root}} \longrightarrow \underset{\delta'=\frac{1}{\pi(\mathbf{p})}\left(1+t'_\pi\right)}{\text{set}} \longrightarrow \underset{\delta'>0}{\text{root}}$$

$$\longrightarrow \underset{\text{as in Def. 2}}{\text{set } \mathbf{p}'=\mathbf{p}'(\delta',\mathbf{x}),\ \mathbf{x}'=\mathbf{x}'(\mathbf{p}')} \longrightarrow \underset{\mathbf{x}'=\mathbf{x},\ \delta'=\frac{1}{\pi(\mathbf{p}')}}{\text{observe}} \tag{2.166}$$

We use the following example to demonstrate the above process in both directions.

Example 2. *For $n = 2$, let $\mathbf{x} = (1/4, 1/5)$. Solving $f(\delta_t, \mathbf{x}) = 0$ gives two positive roots $\delta = (11 - \sqrt{41})/2 \approx 2.29844$, $\delta' = (11 + \sqrt{41})/2 \approx 8.70156$. These two roots are also shown in Fig. 2.2 (the green curve).*

- *If we start from δ, then $\mathbf{p} \equiv \mathbf{p}(\delta, \mathbf{x}) = \left(\frac{1}{40}\left(21 - \sqrt{41}\right), \frac{1}{40}\left(19 - \sqrt{41}\right)\right) \approx (0.364922, 0.314922)$. Solving $g(t_\pi, \mathbf{p}) = 0$ yields the non-zero root as $t'_\pi = \frac{1}{40}\left(41 + 11\sqrt{41}\right) \approx 2.78586$ with which it can be verified that $\delta'_t = \frac{1}{\pi(\mathbf{p})}\left(1 + t'_\pi\right)$ equals δ'. We can also verify $\delta = \frac{1}{\pi(\mathbf{p})}$.*
- *Now if we start from δ', then $\mathbf{p}' \equiv \mathbf{p}(\delta', \mathbf{x}) = \left(\frac{1}{40}\left(21 + \sqrt{41}\right), \frac{1}{40}\left(19 + \sqrt{41}\right)\right) \approx (0.685078, 0.635078)$. Solving $g(t''_\pi, \mathbf{p}') = 0$ yields the non-zero root as $t''_\pi = \frac{1}{40}\left(41 - 11\sqrt{41}\right) \approx -0.735859$ with which it can be verified that $\delta''_t = \frac{1}{\pi(\mathbf{p}')}\left(1 + t''_\pi\right)$ equals δ. We can also verify $\delta' = \frac{1}{\pi(\mathbf{p}')}$.*

2.8.2 Proof of Lemma 6

Since the hyperplane passes through $\mathbf{c}$, the center of the ellipsoid, we can make the translation so that the ellipsoid is centered at the origin and the hyperplane also passes through the origin. Then the ellipsoid has the form $\mathcal{E} : \left\{\mathbf{x} : \mathbf{x}^\mathsf{T}\mathbf{R}^{-1}\mathbf{x} < 1\right\}$, $\mathbf{R} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\mathsf{T}$, where $\mathbf{Q} = [\mathbf{q}_1, \cdots, \mathbf{q}_n]$ is orthonormal and holds the eigenvectors of $\mathbf{R}$, and $\mathbf{D} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ holds the eigenvalues of $\mathbf{R}$. We use $\mathrm{Ref}_{\mathcal{H}(\mathbf{n},d)}(\mathbf{v})$ to denote the reflection of point $\mathbf{v}$ w.r.t. hyperplane $\mathcal{H}(\mathbf{n}, d)$.

"$\Rightarrow$":

The forward part of this lemma requires us to verify:

$$\mathbf{R}\mathbf{q}_i = \lambda_i\mathbf{q}_i, \forall i \in [n] \ \Rightarrow\ \mathrm{Ref}_{\mathcal{H}(\mathbf{q}_i,0)}(\mathbf{v}) \in \partial\mathcal{E} \text{ for all } \mathbf{v} \in \partial\mathcal{E}. \tag{2.167}$$

Since $\mathbf{R}$ is symmetric and has the aforementioned decomposition $\mathbf{R} = \mathbf{Q}\mathbf{D}\mathbf{Q}^\mathsf{T}$, it is easy to verify

that $\mathbf{R}^{-1}$ is symmetric and, if λ_i is an eigenvalue of $\mathbf{R}$ with associated eigenvector $\mathbf{q}_i$ then λ_i^{-1} is an eigenvalue of $\mathbf{R}^{-1}$ with associated eigenvector $\mathbf{q}_i$, namely:

$$\mathbf{R}^{-1}\mathbf{q}_i = \lambda_i^{-1}\mathbf{q}_i. \tag{2.168}$$

Recall the reflecting point w.r.t. $\mathcal{H}(\mathbf{q}_i, 0)$ can be determined by $\mathrm{Ref}_{\mathcal{H}(\mathbf{q}_i,0)}(\mathbf{v}) = \mathbf{v} - 2\frac{\mathbf{v}^\mathsf{T}\mathbf{q}_i}{\mathbf{q}_i^\mathsf{T}\mathbf{q}_i}\mathbf{q}_i$. Therefore $\mathrm{Ref}_{\mathcal{H}(\mathbf{q}_i,0)}(\mathbf{v}) \in \partial\mathcal{E}$ for all $\mathbf{v} \in \partial\mathcal{E}$ means:

$$\left(\mathbf{v} - 2\frac{\mathbf{v}^\mathsf{T}\mathbf{q}_i}{\mathbf{q}_i^\mathsf{T}\mathbf{q}_i}\mathbf{q}_i\right)^\mathsf{T} \mathbf{R}^{-1} \left(\mathbf{v} - 2\frac{\mathbf{v}^\mathsf{T}\mathbf{q}_i}{\mathbf{q}_i^\mathsf{T}\mathbf{q}_i}\mathbf{q}_i\right) = 1, \text{ for all } \mathbf{v} \text{ satisfying } \mathbf{v}^\mathsf{T}\mathbf{R}^{-1}\mathbf{v} = 1, \tag{2.169}$$

which is equivalent to verifying

$$\frac{\mathbf{v}^\mathsf{T}\mathbf{q}_i}{\mathbf{q}_i^\mathsf{T}\mathbf{q}_i}\left(\mathbf{q}_i^\mathsf{T}\mathbf{R}^{-1}\mathbf{v} + \mathbf{v}^\mathsf{T}\mathbf{R}^{-1}\mathbf{q}_i - 2\frac{\mathbf{v}^\mathsf{T}\mathbf{q}_i}{\mathbf{q}_i^\mathsf{T}\mathbf{q}_i}\mathbf{q}_i^\mathsf{T}\mathbf{R}^{-1}\mathbf{q}_i\right) = 0, \text{ for all } \mathbf{v} \in \partial\mathcal{E} \tag{2.170}$$

Applying (2.168) (after transpose), the term in the above parentheses equals $\lambda_i^{-1}\mathbf{q}_i^\mathsf{T}\mathbf{v} + \mathbf{v}^\mathsf{T}\lambda_i^{-1}\mathbf{q}_i - 2\frac{\mathbf{v}^\mathsf{T}\mathbf{q}_i}{\mathbf{q}_i^\mathsf{T}\mathbf{q}_i}\mathbf{q}_i^\mathsf{T}\lambda_i^{-1}\mathbf{q}_i = 0$.

"$\Leftarrow$":

The converse part asks us to show for any vector $\mathbf{q}$ such that $\mathcal{H}(\mathbf{q}, 0)$ is a reflecting hyperplane meaning $\mathrm{Ref}_{\mathcal{H}(\mathbf{q}_i,0)}(\mathbf{v}) \in \partial\mathcal{E}$ for all $\mathbf{v} \in \partial\mathcal{E}$, it has to be an eigenvector of $\mathbf{R}$ meaning there exists some $i \in [n]$ so that $\mathbf{R}\mathbf{q} = \lambda_i\mathbf{q}$.

We prove by contradiction. Suppose for all $i \in [n]$ it holds true that $\mathbf{R}\mathbf{q} \neq \lambda_i\mathbf{q}$, or equivalently $\mathbf{R}^{-1}\mathbf{q} \neq \lambda_i^{-1}\mathbf{q}$.

Recall the qualification (2.169) or equivalently (2.170) holds, which reduces to the following (because we can choose $\mathbf{v}$ such that it is not orthogonal to $\mathbf{q}$: $\mathbf{v}^\mathsf{T}\mathbf{q} \neq 0$) where we also apply the fact that $\mathbf{R}^{-1}$ is symmetric:

$$\left(\mathbf{R}^{-1}\mathbf{q}\right)^\mathsf{T}\mathbf{v} + \mathbf{v}^\mathsf{T}\left(\mathbf{R}^{-1}\mathbf{q}\right) - 2\frac{\mathbf{v}^\mathsf{T}\mathbf{q}}{\mathbf{q}^\mathsf{T}\mathbf{q}}\mathbf{q}^\mathsf{T}\left(\mathbf{R}^{-1}\mathbf{q}\right) = 0. \tag{2.171}$$

On the other hand, the LHS of the above equation can be viewed as a function of $\mathbf{R}^{-1}\mathbf{q}$, which

implies: if $\mathbf{q}$ is not an eigenvector meaning $\mathbf{R}^{-1}\mathbf{q} \neq \lambda_i^{-1}\mathbf{q}$ for all $i \in [n]$, we have:

$$\text{LHS of (2.171)} \neq \left(\lambda_i^{-1}\mathbf{q}\right)^{\mathsf{T}}\mathbf{v} + \mathbf{v}^{\mathsf{T}}\left(\lambda_i^{-1}\mathbf{q}\right) - 2\frac{\mathbf{v}^{\mathsf{T}}\mathbf{q}}{\mathbf{q}^{\mathsf{T}}\mathbf{q}}\mathbf{q}^{\mathsf{T}}\left(\lambda_i^{-1}\mathbf{q}\right) = \lambda_i^{-1}\mathbf{q}^{\mathsf{T}}\mathbf{v} + \mathbf{v}^{\mathsf{T}}\lambda_i^{-1}\mathbf{q} - 2\lambda_i^{-1}\mathbf{v}^{\mathsf{T}}\mathbf{q} = 0, \tag{2.172}$$

which is a contradiction.

2.8.3 Proof of Prop. 17

We classify $\partial\mathcal{S}$ into two categories: $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})| = 1$ (quasi-uniform points) and $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})| > 1$. There are two steps in this proof: Step 1 is to show all the points with $|\mathcal{V}(\mathbf{p})| > 1$ cannot be global extremizers, and Step 2 is to show for points with $|\mathcal{V}(\mathbf{p})| = 1$ the global extremizers are $\{\mathbf{e}_i\}_{i=1}^n$. Finally, it remains to evaluate the objective function at each $\mathbf{e}_i$.

Step 1:

Fix $\mathbf{p}$ such that $|\mathcal{V}(\mathbf{p})| > 1$; we will show such a $\mathbf{p}$ cannot be a global extremizer. Prop. 16 implies that if a potential extremizer has $|\mathcal{V}(\mathbf{p})| > 1$, then it has exactly two distinct non-zero component values namely $|\mathcal{V}(\mathbf{p})| = 2$. Then, suppose $\mathbf{p}$ has n' non-zero components, taking some value $p_s > 0$ for k of those n' components, and some value $p_l > p_s$ for the remaining $n' - k$ components. Note again s stands for small and l for large, and they do not denote indices. Correspondingly $\pi_s = \pi(\mathbf{p})/(1-p_s)$, $\pi_l = \pi(\mathbf{p})/(1-p_l)$. Since $\mathbf{p} \in \partial\mathcal{S}$, we have:

$$kp_s + (n'-k)p_l = 1, \tag{2.173}$$

where $0 < p_s < \frac{1}{n'} < p_l < 1$.

Recall (2.111) in the proof of Prop. 16 needs to be satisfied. Therefore we specialize the first equality in (2.111) to $\mathcal{E}_{\text{out}}$ with such a $\mathbf{p}$ and get:

$$\frac{nc - (kp_s\pi_s + (n'-k)p_l\pi_l)}{(nc-1)c} = \frac{n(\pi_s + p_l\pi_l) - (kp_s\pi_s + (n'-k)p_l\pi_l)}{(n-1)c} \tag{2.174}$$

After some algebra, we have:

$$\begin{aligned} nc(n-1) &= n\pi_s\left(nc-1+kp_s(1-c)\right)+np_l\pi_l\left((n'-k)(1-c)+nc-1\right) \\ c(n-1) &= \frac{\pi(\mathbf{p})}{1-p_s}\left(nc-1+kp_s(1-c)\right)+\frac{p_l}{1-p_l}\pi(\mathbf{p})\left((n'-k)(1-c)+nc-1\right) \\ \frac{c(n-1)(1-p_s)(1-p_l)}{\pi(\mathbf{p})} &= (1-p_l)\left(nc-1+kp_s(1-c)\right)+p_l(1-p_s)\left((n'-k)(1-c)+nc-1\right) \end{aligned}$$

The RHS can be expanded and simplified using (2.173) and becomes $(n-1)(c-p_sp_l)$, so the above equality after canceling $(n-1)$ becomes:

$$\frac{c}{\pi(\mathbf{p})/((1-p_s)(1-p_l))}=c-p_sp_l,$$

which does not hold, because the LHS is greater than c while the RHS is less than c. So we have shown a potential candidate extremizer $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})|>1$ does not exist because of the unsatisfiability of (2.174).

Step 2:

Fix $\mathbf{p}$ such that $|\mathcal{V}(\mathbf{p})|=1$. Recall for all $\mathbf{p}\in\partial\mathcal{S}$ the objective function is initially derived (without any additional assumptions) as $f_{obj}=\frac{1}{na_1^2}\left(nc-\Sigma\right)^2+\frac{1}{a_2^2}\left(S-\frac{1}{n}\Sigma^2\right)$, where $\Sigma=\sum_{i=1}^n x_i$, $S=\sum_{i=1}^n x_i^2$. A point $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})|=1$ is parameterized by an integer $k\in[1,n]$, meaning $\mathbf{p}$ has exactly k non-zero components. Since $\mathbf{p}\in\partial\mathcal{S}$ so each non-zero component equals $1/k$, furthermore, $S(k)=\Sigma(k)^2/k$ and when $k\geq 2$ we can write $\Sigma(k)=\left(1-\frac{1}{k}\right)^{k-1}$.

Observe for both $\mathcal{E}_{\text{in}}$ and $\mathcal{E}_{\text{out}}$, a_2^2 and a_1^2 can be related by:

$$a_2^2=\frac{(n-1)a_1^2}{na_1^2-(nc-1)^2}=\frac{(n-1)a_1^2}{v}, \tag{2.175}$$

where the corresponding denominators (for $\mathcal{E}_{\text{in}}$ and $\mathcal{E}_{\text{out}}$ respectively)

$$v_{\text{in}}\equiv na_{1,\text{in}}^2-(nc-1)^2=(2nc-1-nm)(1-nm),\quad v_{\text{out}}\equiv na_{1,\text{out}}^2-(nc-1)^2=nc-1. \tag{2.176}$$

A scaled version of the objective function can then be rewritten as

$$na_1^2 \cdot f_{obj} = (nc - \Sigma(k))^2 + \frac{n/k - 1}{n-1} \cdot v \cdot \Sigma(k)^2. \tag{2.177}$$

We shall show that for both $\mathcal{E}_{\text{in}}$ (in the proof of Prop. 18) and $\mathcal{E}_{\text{out}}$ (in this proof), the RHS of (2.177) is monotone increasing in k for $k \in [2, n]$ (this assertion is not true for $k \in [1, n]$). Consequently, for $\mathcal{E}_{\text{out}}$ we only need to verify the objective function evaluated at $k = 1$ as well as at $k = 2$ is no smaller than 1, for $\mathcal{E}_{\text{in}}$ we only need to verify the objective function evaluated at $k = 1$ as well as at $k = n$ is no larger than 1.

Towards this, we take derivative of the RHS of (2.177) w.r.t. k and normalize it by $\Sigma(k)^2$, we get

$$-2\left(\frac{nc}{\Sigma(k)} - 1\right) h(k) - \frac{n}{n-1}\frac{1}{k^2}v + 2\frac{n/k - 1}{n-1} v \cdot h(k) \equiv f_d(k), \tag{2.178}$$

where $h(k) \equiv \frac{1}{k} + \log\left(1 - \frac{1}{k}\right) < 0$. Note $\frac{\mathrm{d}}{\mathrm{d}k}\Sigma(k) = \Sigma(k) \cdot h(k)$ for $k \geq 2$. We need to show $f_d(k) \geq 0$.

Recall $\Sigma(k) = \left(1 - \frac{1}{k}\right)^{k-1}$ is monotone decreasing in $k \in [2, \infty)$ from 1/2 to 1/e, it suffices to show a lower bound of $f_d(k)$ is nonnegative, i.e.,

$$f_{dl}(k) \equiv -2\left(2nc - 1\right) h(k) - \frac{n}{n-1}\frac{1}{k^2}v + 2\frac{n/k - 1}{n-1} v \cdot h(k) \geq 0, \tag{2.179}$$

or equivalently,

$$2h(k)\left[-\left(2nc - 1\right) + \frac{n/k - 1}{n-1}v\right] \geq \frac{n}{n-1}\frac{1}{k^2}v. \tag{2.180}$$

It can be verified (by possibly using an upper bound on v) that for both $\mathcal{E}_{\text{in}}$ and $\mathcal{E}_{\text{out}}$ the above bracket term is negative. As will turn out, a convenient and sufficently tight upper bound on $h(k)$ is crucial for analytical tractability. For this, recall the Taylor expansion $\log(1+x) = \sum_{n=1}^{\infty}(-1)^{n-1}\frac{x^n}{n}$ for all $x \in (-1, 1]$. If $x < 0$ then any truncation of this series yields an upper bound. In particular, we have $\log(1 + x) < x - \frac{x^2}{2} + \frac{x^3}{3}$ for $x < 0$. Now setting $x = -\frac{1}{k}$ gives a strict upper bound i.e., $h(k) < -\frac{1}{k^2}\left(\frac{1}{2} + \frac{1}{3k}\right)$. We next show (2.180) for $\mathcal{E}_{\text{out}}$.

Applying the upper bound on $h(k)$ and substituting $v_{\text{out}} = nc - 1$, to show (2.180) it suffices to

verify

$$-2\frac{1}{k^2}\left(\frac{1}{2}+\frac{1}{3k}\right)\left[-\left(2nc-1\right)+\frac{n/k-1}{n-1}\left(nc-1\right)\right]\geq\frac{n}{n-1}\frac{1}{k^2}\left(nc-1\right), \tag{2.181}$$

which can be shown to be equivalent to verify (when $c>1/n$)

$$\left(\frac{1}{2}+\frac{1}{3k}\right)\left(1-\frac{1}{2k}-\frac{1}{2}\frac{c-1}{nc-1}\right)\geq\frac{1}{4}. \tag{2.182}$$

For $k\in[2,n]$, $\frac{1}{2}+\frac{1}{3k}\geq\frac{1}{2}+\frac{1}{3n}>0$, $1-\frac{1}{2k}-\frac{1}{2}\frac{c-1}{nc-1}\geq\frac{3}{4}-\frac{1}{2}\frac{c-1}{nc-1}$. So it suffices to show

$$\left(\frac{1}{2}+\frac{1}{3n}\right)\left(\frac{3}{4}-\frac{1}{2}\frac{c-1}{nc-1}\right)\geq\frac{1}{4}, \tag{2.183}$$

or equivalently $\frac{c(3n^2-4)+3n-2}{24n(nc-1)}\geq 0$, which holds true for $c>1/n$.

So far we have shown for $\mathcal{E}_{\text{out}}$ the objective function expressed as $f_{obj}=\frac{1}{na_1^2}\left(nc-\Sigma\right)^2+\frac{1}{a_2^2}\left(S-\frac{1}{n}\Sigma^2\right)$ is monotone increasing in k for $k\in[2,n]$. Now that for $\mathcal{E}_{\text{out}}$ we want to show the global minimum of the objective function is no smaller than 1, it only requires us to verify $f_{obj}\geq 1$ when $k=1$ and 2. When $k=1$ i.e., at $\mathbf{e}_i$, by construction it is tight so $f_{obj}|_{k=1}=1$. When $k=2$, we need to verify

$$f_{obj}|_{k=2}=\frac{1}{nc(nc-1)}\left(nc-\frac{1}{2}\right)^2+\frac{1}{(n-1)c}\left(\frac{1}{8}-\frac{1}{4}\frac{1}{n}\right)\geq 1. \tag{2.184}$$

The partial derivative w.r.t. c is negative for all $c>1/n$. Since $\lim_{c\to\infty}f_{obj}|_{k=2}=1$, this verifies (2.184).

2.8.4 Proof of Prop. 18

The proof of Λ_{ei} is similar to but more involved than that of Λ_{eo}. In the following we will omit the parts that are common to both proofs. The general strategies are the same. Step 1 is to show all points $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})|>1$ cannot be global extremizers. Step 2 is to show that among the points $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})|=1$, the global extremizers are $\mathbf{e}_i$ as well as $\mathbf{m}$. Finally, it remains to evaluate the

objective function at $\mathbf{e}_i$ or $\mathbf{m}$.

Step 1:

We first recall again a result that is used a few times during the rest of the proof: the function $(1-1/n)^{n-1}$ is monotone decreasing in $n \in [2,\infty)$ from 1/2 to 1/e. Its proof is by taking its derivative and applying the inequality $\log(1+x) \le x$ for all $x > -1$.

Similar to what has been done for $\mathcal{E}_{\text{out}}$. Specializing the first equality in (2.111) to $\mathcal{E}_{\text{in}}$ with such a $\mathbf{p}$ satisfying (2.173), we have:

$$\frac{nc-(kp_s\pi_s+(n'-k)p_l\pi_l)}{a_{1,\text{in}}^2} = \frac{n(\pi_s+p_l\pi_l)-(kp_s\pi_s+(n'-k)p_l\pi_l)}{\frac{(n-1)a_{1,\text{in}}^2}{v_{\text{in}}}},$$

where $v_{\text{in}} = na_{1,\text{in}}^2-(nc-1)^2 = (2nc-1-nm)(1-nm)$. Simplifying using (2.173), we have:

$$\frac{nc(1-p_s)(1-p_l)}{\pi(\mathbf{p})} = 1+v_{\text{in}}-n'p_sp_l-v_{\text{in}}\frac{n-n'}{n-1}p_sp_l. \tag{2.185}$$

In order to show all those $\mathbf{p}$ cannot be global maximizers, we need to show the associated f_{obj} is less than 1. Recall from the proof of Prop. 16 this objective function assuming (2.111) is given in (2.113).

Applying our knowledge of $\mathbf{p}$ and using (2.173), we have

$$f_{obj} = \frac{\frac{\pi(s,l,n')}{(1-p_s)(1-p_l)}}{a_{2,\text{in}}^2}\left(-\pi(s,l,n')+c(n-1)-c(n-n')p_sp_l\right),$$

where $\pi(s,l,n') \equiv (1-p_s)^k(1-p_l)^{n'-k}$ for p_s, p_l satisfying (2.173).

Substituting $a_{2,\text{in}}^2 = \frac{(n-1)a_{1,\text{in}}^2}{v_{\text{in}}}$, $a_{1,\text{in}}^2 = n(c-m)^2$, and (2.185), we see $f_{obj} < 1$ is equivalent to

$$\left(c-\frac{\pi(s,l,n')}{(n-1)\left(1-\frac{n-n'}{n-1}p_sp_l\right)}\right)\left(c-\frac{\pi(s,l,n')}{(1-p_s)(1-p_l)}\left(\frac{1}{n}-\frac{n'}{n}p_sp_l\right)\right) < (c-m)^2. \tag{2.186}$$

Note $c > 1/n$ gives $c > \frac{\pi(s,l,n')}{(1-p_s)(1-p_l)}\left(\frac{1}{n}-\frac{n'}{n}p_sp_l\right)$. We now verify it also ensures $c > \frac{\pi(s,l,n')}{(n-1)\left(1-\frac{n-n'}{n-1}p_sp_l\right)}$. Namely we need to show $\pi(s,l,n') < 1-\frac{1+(n-n')p_sp_l}{n}$, for which we apply AM-GM inequality to

$\pi(s,l,n')$

$$\pi(s,l,n') = (1-p_s)^k(1-p_l)^{n'-k} \leq \left(\frac{k(1-p_s)+(n'-k)(1-p_l)}{n'}\right)^{n'} = \left(1-\frac{1}{n'}\right)^{n'}. \quad (2.187)$$

So it suffices to show

$$\left(1-\frac{1}{n'}\right)^{n'} < 1-\frac{1+(n-n')p_sp_l}{n}. \quad (2.188)$$

When $n'=n$, (2.188) holds true. When $n'<n$, it suffices to show $\frac{1}{n'} > \frac{1+(n-n')p_sp_l}{n}$, or equivalently, $\frac{1}{n'} > p_sp_l$, which holds true as well (as $0<p_s<\frac{1}{n'}<p_l<1$).

Therefore we can apply the AM-GM inequality to the LHS of (2.186), for which it suffices to show:

$$\left(c-\frac{1}{2}\left(\frac{\pi(s,l,n')}{(n-1)\left(1-\frac{n-n'}{n-1}p_sp_l\right)}+\frac{\pi(s,l,n')}{(1-p_s)(1-p_l)}\left(\frac{1}{n}-\frac{n'}{n}p_sp_l\right)\right)\right)^2 < (c-m)^2. \quad (2.189)$$

Clearly $c>m$ too. So next we need to show for those $\mathbf{p}$'s that do satisfy (2.185):

$$f_r(k,p_s,n') \equiv \frac{1}{2}\left(\frac{\pi(s,l,n')}{(n-1)\left(1-\frac{n-n'}{n-1}p_sp_l\right)}+\frac{\pi(s,l,n')}{(1-p_s)(1-p_l)}\left(\frac{1}{n}-\frac{n'}{n}p_sp_l\right)\right)-\frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1} > 0. \quad (2.190)$$

Similar to what has been done in the proof of the spherical inner bound Λ_{si} (Prop. 9), we will show (2.190) by working with an enlarged feasible set, meaning although (2.190) is derived assuming (2.185) (resulted from (2.111) and (2.173)), when we optimize the LHS of (2.190) we do not enforce (2.185), namely the feasible set is enlarged to be all the points $\mathbf{p}$ with $|\mathcal{V}(\mathbf{p})|>1$.

First consider the scenario where $\mathbf{p}$ does not have zero component(s), then the scenario where $\mathbf{p}$ has zero component(s).

Case 1: $\mathbf{p}$ does not have zero component(s), namely $n'=n$.

When $n'=n$, (2.185) simplifies to be

$$\frac{nc}{\pi(s,l,n)/\left((1-p_s)(1-p_l)\right)} = 1+v_{\text{in}}-np_sp_l. \quad (2.191)$$

For fixed $p_s \in (0, 1/n)$, we take the derivate of the denominator of the above LHS w.r.t. k:

$$\frac{\partial}{\partial k}\left(\frac{(1-p_s)^k(1-p_l)^{n-k}}{(1-p_s)(1-p_l)}\right) =$$
$$(1-p_s)^{k-1}\left(1-\frac{1-kp_s}{n-k}\right)^{n-k-1}\left(-\frac{(n-k-1)(1-np_s)}{(n-k)^2\left(1-\frac{1-kp_s}{n-k}\right)}+\log(1-p_s)-\log\left(1-\frac{1-kp_s}{n-k}\right)\right),$$

where p_l is solved from (2.173). Denote $g_1(k,p_s) = -\frac{(n-k-1)(1-np_s)}{(n-k)^2\left(1-\frac{1-kp_s}{n-k}\right)}+\log(1-p_s)-\log\left(1-\frac{1-kp_s}{n-k}\right)$, note $g_1(k,p_s)$ determines the sign of the above derivative. Taking partial derivative of $g_1(k,p_s)$ w.r.t. k, we get

$$\frac{\partial}{\partial k}g_1(k,p_s) = \frac{(np_s-1)(k(n-2)p_s+k-(n-1)(np_s+1))}{(n-k)^2(k(p_s-1)+n-1)^2}, \tag{2.192}$$

which can be verified to be positive for all $k \in (0,n)$ and $p_s \in (0,1/n)$. We then evaluate

$$g_1(0,p_s) = \log\left(1+\frac{\frac{1}{n}-p_s}{1-\frac{1}{n}}\right)-\left(\frac{1}{n}-p_s\right) \overset{(a)}{\geq} \frac{\frac{1}{n}-p_s}{1-\frac{1}{n}}-\frac{1}{2}\left(\frac{\frac{1}{n}-p_s}{1-\frac{1}{n}}\right)^2-\left(\frac{1}{n}-p_s\right) > 0 \text{ for all } n \geq 2, \tag{2.193}$$

where in (a) we apply the inequality $\log(1+x) > x - \frac{1}{2}x^2$ for all $x > 0$. So this implies $g_1(k,p_s)$ itself is positive for all $k \in (0,n)$, and hence this means for fixed $p_s \in (0,1/n)$ the LHS of (2.191) is monotone decreasing in k (as its denominator is increasing in k).

Recall $nm = (1-1/n)^{n-1}$ is monotone decreasing from 1/2 to 1/e, so an upper bound on the RHS of (2.191) is

$$1+\left(2nc-1-\frac{1}{\mathrm{e}}\right)\left(1-\frac{1}{\mathrm{e}}\right)-np_sp_l = \frac{1}{\mathrm{e}^2}-np_sp_l+2\left(1-\frac{1}{\mathrm{e}}\right)nc < \frac{1}{\mathrm{e}^2}+2\left(1-\frac{1}{\mathrm{e}}\right)nc. \tag{2.194}$$

Recall k's actual support consists of integers from $(0,n)$. So in order to show (2.191) is *possibly* satisfiable only when $k = n-1$, due to the monotone decreasing property (w.r.t. k) of its LHS, it suffices to show at $k = n-2$, (2.191) still can not be satisfied, i.e.,

$$\frac{nc}{\pi(s,l,n)/\left((1-p_s)(1-p_l)\right)} > \frac{1}{\mathrm{e}^2}+2\left(1-\frac{1}{\mathrm{e}}\right)nc, \text{ for } k = n-2, \tag{2.195}$$

or equivalently,

$$(1-p_s)^{n-3}\left(\frac{1}{2}+\frac{n-2}{2}p_s\right)=\pi(s,l,n)/\left((1-p_s)(1-p_l)\right)<\frac{nc}{\frac{1}{\mathrm{e}^2}+2\left(1-\frac{1}{\mathrm{e}}\right)nc}. \tag{2.196}$$

The derivative of the LHS w.r.t. p_s is $-\frac{1}{2}(1-p_s)^{n-4}\left(-1+(n-2)^2p_s\right)$, meaning there exists a single stationary point $p_s^*=1/(n-2)^2\in(0,1/n)$ if $n\geq 5$ and in fact this stationary point is the unique maximizer (by verifying the second derivative to be negative at p_s^*). So we will need to show (2.196) holds at p_s^* for $n\geq 5$, i.e.,

$$\left(1-\frac{1}{(n-2)^2}\right)^{n-3}\left(1+\frac{1}{n-2}\right)<\frac{2}{\frac{1}{nce^2}+2\left(1-\frac{1}{\mathrm{e}}\right)}. \tag{2.197}$$

Since $c>1/n$, $n\geq 5$, $1-\frac{1}{(n-2)^2}\in(0,1)$, it suffices to show

$$\left(1-\frac{1}{(n-2)^2}\right)\left(1+\frac{1}{n-2}\right)<\frac{2}{\frac{1}{\mathrm{e}^2}+2\left(1-\frac{1}{\mathrm{e}}\right)}\approx 1.429, \tag{2.198}$$

which can be verified to be true for all $n\geq 5$. For $n=3$ or 4, (2.196) can also be verified to be true (in both cases the LHS maximizes when $p_s=1/n$). $n=2$ case can be skipped since k has to be a positive integer.

Hence (2.191) is possibly satisfiable only when $k=n-1$, for which $\pi(s,l,n)=(1-p_s)^{n-1}(n-1)p_s$ and f_r in (2.190) becomes (for $n'=n$)

$$f_r(k,p_s,n)|_{k=n-1}=\frac{1}{2}(1-p_s)^{n-2}\left((n-2)p_s^2+\frac{1}{n}\right)-\frac{1}{n}\left(1-\frac{1}{n}\right)^{n-1}. \tag{2.199}$$

Since its derivative w.r.t. p_s can be verified to be negative for $p_s\in(0,1/n)$ meaning it is monotone decreasing in p_s, so $f_r(k,p_s,n)|_{k=n-1}>f_r(k,p_s,n)|_{k=n-1,\ p_s=1/n}=0$. This shows (2.190) holds true, and concludes the case when a potential extremizer $\mathbf{p}$ does not have zero component(s).

Case 2: $\mathbf{p}$ has zero component(s), namely $n'<n$.

We first assert that when $n'<n$, it also holds true that (2.185) is satisfiable only by those $\mathbf{p}$

such that $k = n' - 1$. The proof builds upon results from the previous case (i.e., when $n' = n$). Recall there we showed for fixed $p_s \in (0, 1/n)$ the LHS of (2.191) is monotone decreasing in k, and when $k = n - 2$ it is still above the RHS. Note (2.185) can be rewritten as:

$$\frac{n'c}{\pi(s,l,n')/\left((1-p_s)(1-p_l)\right)} = \frac{n'}{n}\left(1 - n'p_sp_l + v_{\text{in}}\left(1 - \frac{n-n'}{n-1}p_sp_l\right)\right). \tag{2.200}$$

Note its LHS is the same as that of (2.191) by reparameterizing n in (2.191) with n', and this allows us to claim that the LHS of (2.200) is monotone decreasing in $k < n'$, and furthermore, it *suffices* to show (2.200)'s RHS is no larger than $\frac{1}{\mathrm{e}^2} + 2\left(1 - \frac{1}{\mathrm{e}}\right)n'c$, because we can reparameterize (2.195) using n'. Namely, we want to show:

$$\frac{n'}{n}\left(1 - n'p_sp_l + v_{\text{in}}\left(1 - \frac{n-n'}{n-1}p_sp_l\right)\right) \le \frac{1}{\mathrm{e}^2} + 2\left(1 - \frac{1}{\mathrm{e}}\right)n'c. \tag{2.201}$$

Applying the upper bound on v_{in} and recognizing $1 - \frac{n-n'}{n-1}p_sp_l \in (0,1)$, it suffices to show:

$$\begin{aligned}\frac{n'}{n}\left(1 - n'p_sp_l + \left(2nc - 1 - \frac{1}{\mathrm{e}}\right)\left(1 - \frac{1}{\mathrm{e}}\right)\right) &\le \frac{1}{\mathrm{e}^2} + 2\left(1 - \frac{1}{\mathrm{e}}\right)n'c \\ \frac{n'}{n}\left(\frac{1}{\mathrm{e}^2} - n'p_sp_l\right) &\le \frac{1}{\mathrm{e}^2},\end{aligned}$$

which is true. Thus we have shown (2.200) is possibly satisfiable only when $k = n' - 1$.

So we investigate f_r when $k = n' - 1$ (or equivalently $n' = k + 1$). In the following we show $f_r(k, p_s, n')|_{n'=k+1}$ is monotone decreasing in p_s for fixed k. Taking derivative

$$\frac{\partial}{\partial p_s} f_r(k, p_s, k+1) = -\frac{k(1-p_s)^{k-2}\left((k+1)p_s^2(-k+n-1) + 3p_s(k-n+1) + n - 2\right)}{2n\left(kp_s^2(-k+n-1) + p_s(k-n+1) + n - 1\right)^2} \cdot g_2(k, p_s),$$

where $g_2(k,p_s) \equiv n\left(p_s\left(k\left(p_s\left((k+1)kp_s^2 - 2(k+1)p_s + k + 4\right) - 2\right) + p_s - 2\right) + 1\right) - ((k+1)p_s(kp_s - 1) + 1)^2$. Since one can show for $0 < p_s < \frac{1}{n'} < 1$ and $k = n' - 1$ and $n' < n$ and $n > 2$ (note if $n = 2$ there is no need to consider $n' < n$ case, as the minimum of possible value for k is 1 and k

also needs to satisfy $k = n' - 1 < n - 1$ strictly), we have

$$(k+1)p_s^2(-k+n-1) + 3p_s(k-n+1) + n - 2 \geq 0.$$

We then need to show the positiveness of $g_2(k, p_s)$. We rearrange terms viewing $g_2(k, p_s)$ as a polynomial of p_s and split the coefficient of the quadratic term:

$$\begin{aligned} g_2(k,p_s) &= \left(k^2(k+1)(n-k-1)p_s^4 - 2k(k+1)(n-k-1)p_s^3 + (k+1)(n-k-1)p_s^2\right) + \\ &\quad \left((nk(k+3) - 2k(k+1))p_s^2 - 2(k+1)(n-1)p_s + n - 1\right) \\ &= (k+1)(n-k-1)p_s^2\left(kp_s - 1\right)^2 + \left((nk(k+3) - 2k(k+1))p_s^2 - 2(k+1)(n-1)p_s + n - 1\right). \end{aligned}$$

It then suffices to show $(nk(k+3) - 2k(k+1))p_s^2 - 2(k+1)(n-1)p_s + n - 1 \geq 0$, which can be verified to be true, since its discriminant is always negative and the coefficient of p_s^2 is positive.

Now showing (2.190) amounts to showing $f_r(k, p_s, n')|_{n'=k+1,\ p_s=1/n'} > 0$.

First we show it is monotone decreasing in k. Towards this, we take derivative:

$$\frac{\mathrm{d}}{\mathrm{d}k} f_r(k, 1/(k+1), k+1) = \frac{k^k(k+1)^{-k-1}}{2n(-(k+2)n + k + 1)^2} \cdot g_3(k), \tag{2.202}$$

where $g_3(k) \equiv (2(k+4)k+7)\,n^2 - (3k+5)(k+1)n + (k+1)(k(n-1)+2n-1)(k(2n-1)+3n-1)\log\left(\frac{k}{k+1}\right) + (k+1)^2$. We need to show $g_3(k)$ is negative, for which it suffices to show (using the inequality $\log(1+x) < x - \frac{1}{2}x^2$, $\forall x \in (-1, 0)$) a *strict* upper bound of it is nonpositive. e.g.,

$$\begin{aligned} &(2(k+4)k+7)n^2 - (3k+5)(k+1)n + \\ &(k+1)(k(n-1)+2n-1)(k(2n-1)+3n-1)\left(-\frac{1}{k+1} - \frac{1}{2}\left(-\frac{1}{k+1}\right)^2\right) + (k+1)^2 \leq 0. \end{aligned}$$

The above LHS can be simplified as $\frac{(k-(n-1))(k(3n-1)+4n-1)}{2(k+1)}$, which is nonpositive for all positive $k \leq n-1$.

Second, since $f_r(k, p_s, n')|_{n'=k+1,\ p_s=1/n'}$ is monotone decreasing in k for all positive $k \leq n-1$,

it is lower bounded by $f_r(k, p_s, n')|_{n'=k+1,\ p_s=1/n',\ k=n-1} = 0$. Therefore, we have shown the desired inequality (2.190) holds true, and this concludes the case when a potential extremizer $\mathbf{p}$ has zero component(s).

Step 2: This step follows exactly Step 2 of the proof of Λ_{eo}, up to the paragraph right below (2.180). Applying the upper bound on $h(k)$, to show (2.180) it suffices to show

$$-2\frac{1}{k^2}\left(\frac{1}{2}+\frac{1}{3k}\right)\left[-(2nc-1)+\frac{n/k-1}{n-1}v_{\text{in}}\right] \geq \frac{n}{n-1}\frac{1}{k^2}v_{\text{in}}, \tag{2.203}$$

which after rearranging terms becomes:

$$\left(1+\frac{2}{3k}\right)(n-1)(2nc-1) \geq \left(n+\left(1+\frac{2}{3k}\right)\left(\frac{n}{k}-1\right)\right)v_{\text{in}}, \tag{2.204}$$

for which we apply an upper bond on v_{in}, multiply both sides by k^2, and seek to show

$$k^2\left(1+\frac{2}{3k}\right)(n-1)(2nc-1) - k^2\left(n+\left(1+\frac{2}{3k}\right)\left(\frac{n}{k}-1\right)\right)\left(2nc-1-\frac{1}{\mathrm{e}}\right)\left(1-\frac{1}{\mathrm{e}}\right) \geq 0. \tag{2.205}$$

Denote the above LHS by $g_4(k)$, to show $g_4(k) \geq 0$ for all $k \in [2, n]$ it suffices to show $\frac{\mathrm{d}}{\mathrm{d}k}g_4(k) \geq 0$ and $g_4(k)|_{k=2} \geq 0$.

$$\frac{\mathrm{d}}{\mathrm{d}k}g_4(k) = \frac{6k(n-1)(2ecn-1) + n\left(-2ec((e-3)n+2)+e^2-3\right)+2}{3e^2}. \tag{2.206}$$

The numerator can be rewritten as

$$12\mathrm{e}ckn^2 - 12\mathrm{e}ckn - 2\mathrm{e}^2cn^2 + 6\mathrm{e}cn^2 - 4\mathrm{e}cn - 6kn + 6k + \mathrm{e}^2n - 3n + 2,$$

or equivalently,

$$6(n-1)(2\mathrm{e}cn-1)k - 2\mathrm{e}^2cn^2 + 6\mathrm{e}cn^2 - 4\mathrm{e}cn + \mathrm{e}^2n - 3n + 2,$$

which viewed as a linear function of k is increasing in k, so to show $\frac{\mathrm{d}}{\mathrm{d}k}g_4(k) \geq 0$ we need to show it is positive when $k = 2$, i.e., $-2\mathrm{e}^2cn^2 + 30\mathrm{e}cn^2 - 28\mathrm{e}cn + \mathrm{e}^2n - 15n + 14 \geq 0$. Since the coefficient of c is $-28\mathrm{e}n + 30\mathrm{e}n^2 - 2\mathrm{e}^2n^2 > 0$ namely this function is increasing in c, it then suffices to show it is positive when c is the minimum possible (i.e., $c = 1/n$), namely, $(30\mathrm{e} - 15 - \mathrm{e}^2)n + 14(1 - 2\mathrm{e}) \geq 0$, which is true for $n \geq 2$.

Having shown $g_4(k)$ is monotone increasing in k, we now only need to show $g_4(k)|_{k=2} \geq 0$, i.e.,

$$-\frac{4\left(2(\mathrm{e}-5)\mathrm{e}cn^2 + \left(8\mathrm{e}c - \mathrm{e}^2 + 5\right)n - 4\right)}{3\mathrm{e}^2} \geq 0,$$

or equivalently,

$$2(\mathrm{e}-5)\mathrm{e}cn^2 + \left(8\mathrm{e}c - \mathrm{e}^2 + 5\right)n - 4 \leq 0. \tag{2.207}$$

Since for fixed n the LHS is monotone decreasing in c (we can check its derivative w.r.t. c equals $2\mathrm{e}n(4 - (5 - \mathrm{e})n) < 0$ for $n \geq 2$). It suffices to show (2.207) holds when c is the minimum possible (i.e., $c = 1/n$). Setting $c = 1/n$ in (2.207), after some algebra it gives $\frac{4(2\mathrm{e}-1)}{10\mathrm{e}-\mathrm{e}^2-5} < n$ which holds true for $n \geq 2$, since $\frac{4(2\mathrm{e}-1)}{10\mathrm{e}-\mathrm{e}^2-5} \approx 1.199577$.

Thus we have shown the desired inequality (2.180) for $\mathcal{E}_{\text{in}}$.

So far we have shown for $\mathcal{E}_{\text{in}}$ the objective function expressed as $f_{obj} = \frac{1}{na_1^2}\left(nc - \Sigma\right)^2 + \frac{1}{a_2^2}\left(S - \frac{1}{n}\Sigma^2\right)$ is monotone increasing in k for $k \in [2, n]$. Now that for $\mathcal{E}_{\text{in}}$ we need to show the global maximum of the objective function is no larger than 1, we only need to verify $f_{obj} \leq 1$ when $k = 1$ and n, both of which are true as by construction the objective function is tight at $\mathbf{e}_i$ (i.e., $k = 1$) and $\mathbf{m}$ (i.e., $k = n$).

Chapter 3: Delay on broadcast erasure channels under random linear combinations

3.1 Introduction

The focus of this chapter is on the number of time slots required to broadcast a collection of c packets to n receivers, where each transmitter to receiver channel is an independent erasure channel with reception probabilities $\mathbf{q} = (q_1, \ldots, q_n)$ (Fig. 3.1). In particular, the random delay associated with the transmission is the number of time slots until each receiver has all c packets, when the transmitter forms random linear combinations of the c packets over a field of size d, denoted $Y_{n:n}^{(c)}$. The focus of our investigation is primarily, although not exclusively, on deriving properties (exact expressions, lower and upper bounds, asymptotics) of the r^{th} moment of $Y_{n:n}^{(c)}$, denoted $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$. The proof techniques we employ are almost entirely probabilistic.

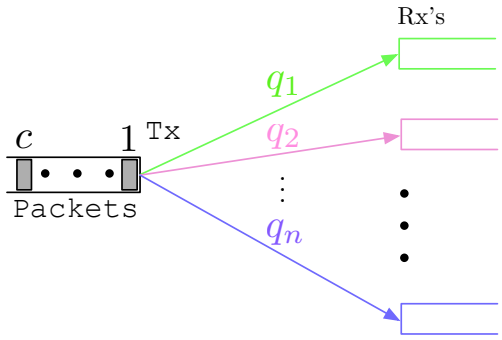

Figure 3.1: The heterogeneous broadcast erasure channel consists of a transmitter and n receivers connected over independent erasure channels with reception probabilities $\mathbf{q} = (q_1, \ldots, q_n)$. Of interest is the time required for all n receivers to receive all c packets, denoted $Y_{n:n}^{(c)}$. The homogeneous channel has $q_j = q$ for $j \in [n]$.

3.1.1 Motivation and related work

The broad motivation and context for this work is the fact that the use of random linear combinations of packets as a coding paradigm has received a great deal of attention within both the network coding (see, e.g.,[31] and subsequent work) and fountain coding (see, e.g.,[32] and subsequent work) communities. More specifically, there have been a number of recent (since 2006) works focused on

the broadcast delay of random linear combinations over erasure channels, including[1;33–37]. We now relate the contributions of this chapter within the context of this body of work. Several of our results are new, some results are extensions of existing results, and some results are new proofs of known results using new (probabilistic) proof techniques.

To our knowledge the earliest work on broadcast delay using random linear combinations over erasure channels is that of Eryilmaz, Ozdaglar, Médard, and Ahmed[1] (with a preliminary conference version in[38]). They compare the delays under scheduling with channel state information vs. random linear coding and establish the superiority of the latter, and then establish explicit expressions for the expected delay under random linear coding, along with asymptotic expressions in the blocklength and number of receivers. We extend and reprove via alternate techniques a couple of their results.

The work of Cogill, Shrader, and Ephremides[33], and Cogill and Shrader[34;35] address (variously) throughput, delay, and stability of multicast queueing systems over erasure channels using random linear packet coding. Of these, the work closest to ours is[33]. Their focus in this work is primarily on the stability region of arrival rates for multicast queueing systems, and in establishing their results they obtain several results on the broadcast delay. Again, we extend and reprove via alternate techniques a couple of their results.

Recent work by Yang and Shroff[36] has extended the channel model to the Markov-modulated erasure channel (allowing correlations in time). Additionally, Swapna, Eryilmaz, and Shroff[37] have studied extensions of this framework to address the case where the blocklength c scales with n, the number of receivers, and in particular they show the existence of a phase-transition at $c(n) = \Theta(\log n)$. Our work does not address either of these extensions — we restrict our attention to erasure channel realizations that are independent in time (and across users), and we do not address the asymptotic regime when the blocklength c grows with the number of users n.

Besides our contributions extending certain results in the above work, we have also investigated the following additional topics that, to our knowledge, have not been previously addressed. First, a common theme throughout our work is on lower and upper bounds on each of the moments of the delay, which hold for all finite c, n, and which we additionally show to be (almost) asymptotically

tight for fixed n as c grows large, and for fixed c as n grows large. Second, we establish the intuitive fact that the expected delay per packet is decreasing in the blocklength c. Although this fact is intuitive, the proof is non-trivial, and in fact we show some (perhaps) counter-intuitive results giving necessary and sufficient conditions on the sample-path realizations of the delay per packet to be decreasing as the blocklength is increased. Third, we employ extreme value theory and stochastic ordering in studying the asymptotic behavior of the delay as the number of receivers n is increased. As will become evident, it is natural to use these two tools together.

3.1.2 Outline

This chapter is structured as follows. The model and common notation are introduced in §3.2, and §3.3 analyzes the delay without coding, i.e., when each packet in the block is repeatedly broadcast until received by all receivers. The next three sections form the heart of the chapter. First, in §3.4 we address the delay under random linear combinations, with subsections for *i*) lower and upper bounds, *ii*) exact expressions, *iii*) a recurrence for the delay, and *iv*) a characterization of the channels that minimize the delay. Second, in §3.5 we address the delay per packet as a function of the blocklength c, with subsections for *i*) the asymptotic (in c) delay per packet, *ii*) monotonicity (in c) properties of the delay per packet, and *iii*) bounds on the expected delay per packet that are asymptotically tight in c. Third, in §3.6 we address the delay as a function of the number of receivers n, where our approach couples order statistic inequalities with stochastic ordering and extreme value theory to establish that the bounds on delay are asymptotically tight in n. Several of the longer proofs are found in the Appendix. Table 3.1 contains a summary of the results in the chapter.

Table 3.1: Summary of results

Section/Result	Title/Description
§3.3	**Delay under uncoded transmission**
§3.3.1	Lower and upper bounds on moments of the delay under UT
Prop. 3	Bounds on $\mathbb{E}[X_{n:n}^r]$ (A2)
Cor. 1	Bounds on $\mathbb{E}[X_{n:n}^r]$ (A3)
§3.3.2	An exact moment expression for the delay under UT
Prop. 4	Exact $\mathbb{E}[X_{n:n}^r]$ (A2)
Cor. 2	Exact $\mathbb{E}[X_{n:n}]$ and $\mathbb{E}[X_{n:n}^2]$ (A2)
Prop. 5	Exact $\mathbb{E}[X^r]$
§3.4	**Delay under random linear combinations**
§3.4.1	Lower and upper bounds on moments of the delay under RLC
Prop. 6	Bounds on $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ (A1)
Prop. 7	Bounds on $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ (A2)
Prop. 8	Bounds on $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ (A3)
§3.4.2	Exact expressions of the moments of delay under RLC
Prop. 9	Exact $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ (A1,A2)
§3.4.3	Recurrence for the (moments of) delay under RLC
Prop. 10	Recurrence for $Y_{n:n}^{(\mathbf{c})}$ (A1,A2)
Cor. 3	Recurrence for $\mathbb{E}[Y_{n:n}^{(\mathbf{c})}]$ (A1,A2)
§3.4.4	Erasure channels that minimize delay
Prop. 11	Minimization of $\mathbb{E}[X_{n:n}^r]$ (A2)
Prop. 12	Minimization of $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ (A2)
§3.5	**Dependence of RLC delay on the blocklength c**
§3.5.1	Asymptotic delay per packet as $c \to \infty$
Prop. 13	Convergence (in c) of $Y_{n:n}^{(c)}/c$ in probability and in r^{th} mean (A1)
Prop. 14	Satisfying conditions for Prop. 13 (A1)
Prop. 15	Inequality $\lim_{c\to\infty} \mathbb{E}[Y_{n:n}^{(c)}]/c \leq \mathbb{E}[X_{n:n}]$ (A1)
§3.5.2	Monotonicity properties of delay per packet
Prop. 16	$\mathbb{E}[Y_{n:n}^{(c)}]/c$ monotone decreasing in c (A2)
Prop. 17	$\mathbb{E}[Y_{n:n}^{(c)}]/c$ monotone decreasing in c (A1)
Cor. 4	Optimal sum block delay over block partitions (A1)
Prop. 18	$Y_j^{(c)}/c$ and $Y_j^{(c+1)}/(c+1)$ not stochastically ordered (A2)
Lem. 1	Sample path $T_{n:n}^{(c',m)}(\omega) \leq T_{n:n}^{(c,m)}(\omega)$ for $m \leq c'$ (A2)
Prop. 19	Sample path $T_{n:n}^{(c',m)}(\omega) \leq T_{n:n}^{(c,m)}(\omega)$ for $m > c'$ (A2)
§3.5.3	Bounds on expected delay per packet
Prop. 20	Asymptotically (in c) tight bounds on $\mathbb{E}[Y_{n:n}^{(c)}]/c$ (A3)
§3.6	**Dependence of RLC delay on the number of receivers n**
§3.6.1	Asymptotic delay as $n \to \infty$
Cor. 5	Bounds on $\lim_{n\to\infty} \mathbb{E}\left[\left(\phi(q)Y_{n:n}^{(c)} - b_n\right)^r\right]$ (A3)
Lem. 2	Standardized and non-standardized asymptotic moments
Prop. 23	Asymptotic (in n) scaling of r^{th} moment: $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ (A3)
§3.6.2	Bounds on moments of delay as $n \to \infty$
Prop. 24	Asymptotically (in n) tight lower bound on $\mathbb{E}[\left(\tilde{Y}_{n:n}^{(c)}\right)^r]$ (A3)
Prop. 25	Asymptotically (in n) tight upper bound on $\mathbb{E}[\left(\tilde{Y}_{n:n}^{(c)}\right)^r]$ (A3)

Table 3.2: Summary of results (cont'd)

Section/Result	Title/Description
Appendices	
§3.7.1	Properties of regularized incomplete Beta function $I_x(a,b)$
Lem. 3	$I_x(a,b)$ is increasing in b
Lem. 4	$\log I_x(a,b)$ is concave on $x \in (0,1)$
§3.7.2	Proof of Prop. 14 (A1)
§3.7.3	Proof of Prop. 16 (monotonicity for infinite field size) (A2)
Lem. 5	Column invariant row-index selection rule (A2)
Lem. 6	$\mathbb{E}\left[\max_{j\in[n]} \sum_{k=1}^{c} Z_{j,k}^{(c+1)}\right] > \mathbb{E}\left[\sum_{k=1}^{c} Z_{\hat{J}^{(c+1)},k}^{(c+1)}\right]$ (A2)
§3.7.4	Proof of Prop. 17 (monotonicity for finite field size) (A1)
Lem. 7	RVs $(\hat{J}, X_{j,k} + X_{j,k'}, X_{j,k}, X_{j,k'})$ form Markov chains (A1)
Lem. 8	Stochastic ordering of geometric RVs (A1)
§3.7.5	Proof of Prop. 19 (A2)
§3.7.6	Lower and upper bounds on the expected maximum order statistic
Prop. 28	Lower and upper bounds for exponential RVs

3.2 Model and notation

In this section we introduce the model and the notation. Relevant discrete distributions are covered in §3.2.1 and continuous distributions in §3.2.2. In many situations (e.g., bounding a moment of the random delay) it is more convenient to work with the "associated" continuous random variable (RV), where the association is through the existence of a stochastic ordering, as will be described below.

Each of the c packets is assumed to be of equal size, each of the n receivers is assumed to want all c packets, and each receiver is assumed to have none of the packets in its possession at the beginning of the transmissions. Let time be slotted with the constant slot duration equal to the transmission time of a packet. We assume a block-fading erasure channel model wherein the channel between the transmitter and receiver j in time slot t is a Bernoulli RV, say $Z_{j,t}$, where $Z_{j,t} = 0$ (1) indicates an erasure (reception), respectively, with $\mathbb{P}(Z_{j,t} = 1) = q_j$ for $j \in [n]$. The RVs $(Z_{j,t},\ j \in [n], t \in \mathbb{N})$ are assumed to be independent in time and across receivers. The transmitter is blind in that the transmitter's actions are independent of the "state" of the receivers, modulo the fact that the transmitter knows to move on to the next block (if one exists) once all the receivers have decoded the block. A problem instance is specified by the tuple $(c, n, \mathbf{q})$, for $c \in \mathbb{N}$, $n \in \mathbb{N}$,

and $\mathbf{q} = (q_j, j \in [n]) \in (0,1)^n$. Table 3.3 lists general notation, while Table 3.4 lists notation for the specific distributions in the chapter.

Our notational convention is to write $Z \sim F_Z$ to indicate the RV Z has a cumulative distribution function (CDF) F_Z. Similarly, $Z_1 \sim Z_2$ and $Z_1 \stackrel{\mathrm{d}}{=} Z_2$ for RVs Z_1, Z_2 means they are equal in distribution.

Table 3.3: General notation (for Chapter 3)

Symbol	Meaning
n, $[n] = \{1, \ldots, n\}$	number and set of receivers
$[n]_s$	set of all subsets of $[n]$ of size s
c, $[c] = \{1, \ldots, c\}$	blocklength and set of packets per block
$\mathbf{q} = (q_j, j \in [n])$	reception probabilities under Assumptions A1, A2
$q_j = q$	reception probabilities for each $j \in [n]$ under Assumption A3
$\mathbf{Q} = (q_{j,k}, j \in [n], k \in [c])$	success probability matrix under Assumption A1
$q_{j,k} = q_j$	success probabilities for each $j \in [n]$ and $k \in [c]$ under Assumption A2
$q_{j,k} = q$	success probabilities for each $j \in [n]$ and $k \in [c]$ under Assumption A3
$\phi(q) = -\log(1-q)$	"rate" parameter of the continuous analog RV
d	field size for coefficients used in linear combinations
$H_n = 1 + 1/2 + \cdots 1/n$	n^{th} harmonic number
$\binom{m}{k}$	binomial coefficient
$\left\{ {m \atop k} \right\}$	Stirling number of the second kind
$Z, \tilde{Z}$	a generic discrete/continuous RV
$Z_{n:n}, \tilde{Z}_{n:n}$	maximum of n independent RVs $(Z_1, \ldots, Z_n)$, $(\tilde{Z}_1, \ldots, \tilde{Z}_n)$
$F_Z(z), F_{\tilde{Z}}(z)$	cumulative distribution function (CDF) for $Z, \tilde{Z}$
$p_Z(z), f_{\tilde{Z}}(z)$	probability mass/density function (PMF/PDF) for $Z, \tilde{Z}$
$\mathbb{E}[Z^r], \mathbb{E}[\tilde{Z}^r]$	r^{th} moment of $Z, \tilde{Z}$

3.2.1 Discrete distributions

Index the transmission slots by $\mathbb{N}$, and define the (random) block delay to be the number of slots until each of the n receivers has successfully decoded the block of c packets. We consider two separate transmission schemes: *i*) uncoded transmission (§3.3), abbreviated UT, and *ii*) random linear combinations (§3.4 through §3.6), abbreviated RLC.

Under UT, the transmitter in effect treats each of the c packets as a separate block, and repeatedly (re-)transmits each packet until all receivers have that packet, at which time it moves on to the next packet in the block of c packets, if any. The presumption is that a feedback channel exists which alerts the transmitter that all n receivers have the packet. Due to the standing independence assumptions, it is clear that the overall random delay to transmit all c packets is the sum of c independent and

Table 3.4: Probability distributions

Symbol	Meaning
$X \sim \text{Geo}(q)$	geometric RV with success probability q
$\tilde{X} \sim \text{Exp}(1)$	exponential RV with rate 1
$\frac{1}{\phi}\tilde{X} \sim \text{Exp}(\phi)$	exponential RV with rate ϕ
$Y_j^{(c)} = X_{j,1} + \cdots + X_{j,c}$	random time required for c successes at receiver j
$Y_j^{(c)} \sim \text{GenNegBin}(c, \mathbf{q}_j)$	$Y_j^{(c)}$ is a generalized negative binomial RV under Assumption A1
$Y_j^{(c)} \sim \text{NegBin}(c, q_j)$	$Y_j^{(c)}$ is a negative binomial RV under Assumption A2
$Y_j^{(c)} \sim \text{NegBin}(c, q)$	$(Y_j^{(c)}, j \in [n])$ are iid negative binomial RVs under Assumption A3
$\tilde{Y}^{(c)} \sim \text{Gamma}(c, 1)$	Gamma RV for sum of c independent unit rate exponential RVs
$\frac{1}{\phi}\tilde{Y}^{(c)} \sim \text{Gamma}(c, \phi)$	Gamma RV for sum of c independent exponential RVs with rate ϕ
$\tilde{Y}_j^{(c)} = \tilde{X}_{j,1} + \cdots + \tilde{X}_{j,c}$	"continuous" analog of $Y_j^{(c)}$
$\tilde{Y}_j^{(c)} \sim \text{HypoExp}(c, \phi(\mathbf{q}_j))$	$\tilde{Y}_j^{(c)}$ is a hypo-exponential RV under Assumption A1
$\tilde{Y}_j^{(c)} \sim \text{Gamma}(c, \phi(q_j))$	$\tilde{Y}_j^{(c)}$ is a Gamma RV under Assumption A2
$\tilde{Y}_j^{(c)} \sim \text{Gamma}(c, \phi(q))$	$(\tilde{Y}_j^{(c)}, j \in [n])$ are iid Gamma RVs under Assumption A3
$W \sim \text{Bin}(m, q)$	binomial RV with m trials and success probability q
$I_x(\alpha, \beta)$	regularized incomplete beta function
$I_q(l, m-l+1) = \mathbb{P}(W \geq l)$	CCDF for W using regularized incomplete Beta function

identically distributed (iid) RVs, each representing the time (in slots) for the transmitter to complete one of the packets. In particular, define the random delay per packet under UT as the maximum of n geometric RVs, $X_{n:n} \equiv \max(X_1, \ldots, X_n)$, with $(X_1, \ldots, X_n)$ independent and $X_j \sim \text{Geo}(q_j)$ representing the delay per packet under UT for receiver j, i.e., the number of transmission attempts until receiver j is successful.

Under RLC, the transmitter forms in each time slot a new random linear combination of the c packets, with coefficients generated uniformly at random from a finite or infinite field of size d, namely $\{0, \ldots, d-1\}$. Each combination of packets is then broadcast to the receivers, and a receiver is able to decode all c packets in the block once it has received any c linearly independent combinations (say, by inverting the corresponding full-rank submatrix of combination coefficients). The block delay under RLC is the number of time slots until all receivers have decoded all c packets.

More precisely, the block delay for a block of c packets under RLC is denoted by $Y_{n:n}^{(c)} \equiv \max\left(Y_1^{(c)}, \ldots, Y_n^{(c)}\right)$. The expected block delay is $\mathbb{E}\left[Y_{n:n}^{(c)}\right]$ and the expected delay per packet is $\mathbb{E}\left[Y_{n:n}^{(c)}\right]/c$. Each $Y_j^{(c)} = \sum_{k=1}^{c} X_{j,k}$ is a generalized negative binomial, denoted $Y_j^{(c)} \sim \text{GenNegBin}(c, \mathbf{q}_j)$ with parameter $\mathbf{q}_j = (q_{j,1}, \ldots, q_{j,c})$, for independent geometric RVs $(X_{j,1}, \ldots, X_{j,c})$ with $X_{j,k} \sim$

Geo($q_{j,k}$). We say that a receiver j is in state $k \in [c]$ if it has already received $k-1$ linearly independent combinations from the transmitter. Then, $X_{j,k}$ represents the time between receiver j obtaining the $(k-1)^{\text{st}}$ and k^{th} linearly independent combination. We reiterate that k is not a time index, per se, although $k = k(t)$ is nondecreasing in t. That $(X_{j,k})$ are independent in j (k) is due to the assumption that the erasure channels are independent in space (time), respectively. The $n \times c$ matrix $\mathbf{Q}$ with entries $q_{j,k}$ for $j \in [n]$ and $k \in [c]$ holds the success probability indexed by receiver j at state k:

$$q_{j,k} = (1 - d^{k-1-c})q_j, \quad j \in [n], \ k \in [c]. \tag{3.1}$$

Here, q_j is the time-invariant channel reception probability for receiver j, d is the field size, and $1-d^{k-1-c}$ is the probability that the linear combination sent by the transmitter is in fact independent of the $k-1$ linear combinations already received by receiver j (e.g.,[33] Lemma 1, p. 275). Eq. (3.1) is the most general case, which we occasionally specialize for tractability, as indicated below.

Assumption 1. *Throughout, we assume one of the three cases listed below, in order of decreasing generality.*

A1: **State-dependent receptions, heterogeneous receivers.** *The field size d is finite and the reception probabilities $\mathbf{q} = (q_1, \ldots, q_n)$ are unrestricted (heterogeneous). The success probabilities are given by the $n \times c$ matrix $\mathbf{Q}$ with entries $q_{j,k}$ in* (3.1)*. The random block delay $Y_{n:n}^{(c)}$ is the maximum of n independent generalized negative binomial RVs $Y_j^{(c)} \sim \text{GenNegBin}(c, \mathbf{q}_j)$, with $\mathbf{q}_j = (q_{j,1}, \ldots, q_{j,c})$.*

A2: **State-independent receptions, heterogeneous receivers.** *The field size d is infinite, but the reception probabilities $\mathbf{q} = (q_1, \ldots, q_n)$ are unrestricted (heterogeneous). Due to the infinite field size, received linear combinations are linearly independent of all previous combinations, and so the success probabilities equal the reception probabilities, which are independent of k, i.e., $q_{j,k} = q_j$ for $j \in [n]$ and $k \in [c]$. The success probabilities are given by $\mathbf{q} = (q_j, j \in [n])$. The random block delay $Y_{n:n}^{(c)}$ is the maximum of n independent negative binomial RVs $Y_j^{(c)} \sim \text{NegBin}(c, q_j)$.*

A3: **State-independent receptions, homogeneous receivers.** *The field size d is infinite, and the reception probabilities are homogeneous, i.e., $q_j = q$ for $j \in [n]$. Due to the infinite field size, received linear combinations are linearly independent of all previous combinations, and so the success probabilities equal the reception probability, which is independent of k, i.e., $q_{j,k} = q$ for each $j \in [n]$ and $k \in [c]$. The random block delay $Y_{n:n}^{(c)}$ is the maximum of n iid negative binomial RVs $Y_j^{(c)} \sim \text{NegBin}(c, q)$.*

In each result, subsection, and section we will indicate the governing assumption. To reiterate, A1 is the most general assumption, A2 is a special case of A1 for $d = \infty$, and A3 is a special case of A2 for $q_j = q$. The governing assumption for each result in the chapter is also listed in Table 3.1.

It is worth noting that, in general, setting $c = 1$ in RLC does not recover UT as a special case. In particular, for $d < \infty$, setting $c = 1$ gives $q_{j,1} = (1 - d^{-1})q_j \neq q_j$, since we are forming linear combinations over a block of one packet, for which there is a probability of selecting the 0 coefficient, which is (trivially) linearly dependent upon previous receptions. Naturally, we *do* recover UT as a special case of RLC for $c = 1$ and $d = \infty$.

3.2.2 Continuous distributions

Although the (discrete) geometric and negative binomial distributions directly capture the discrete delay of interest to us, nonetheless we will often find it useful to consider what we call the *continuous analogs* of these distributions. The lynchpin connecting the discrete RVs to their continuous RV analogs is the notion of stochastic ordering, the basic concepts of which we briefly review below, drawing directly from Ross[39] (Chapter 9). As described below, stochastic ordering is preserved under positive scaling, translation, and component-wise nondecreasing functions (with independent components) and more importantly implies moment ordering. Collectively, these properties allow us to establish inequalities on the r^{th} moment of a discrete RV in terms of the r^{th} moment of its (often more tractable) continuous analog.

Generic scalar (continuous or discrete) RVs Z_1, Z_2 are said to be stochastically ordered, denoted $Z_1 \leq_{\text{st}} Z_2$, if for all z: $\bar{F}_{Z_1}(z) \leq \bar{F}_{Z_2}(z)$ for $\bar{F} = 1 - F$ the complementary CDF (CCDF) of the RV. Stochastic ordering implies moment ordering, i.e., $Z_1 \leq_{\text{st}} Z_2$ implies $\mathbb{E}[Z_1] \leq \mathbb{E}[Z_2]$ ([39],

Lemma 9.1.1). If we instead let Z_1, Z_2 denote random n-vectors with *independent* stochastically ordered components, i.e., $Z_1 = (Z_{1,1}, \ldots, Z_{1,n})$ and $Z_2 = (Z_{2,1}, \ldots, Z_{2,n})$ each have independent components and $Z_{1,j} \leq_{\text{st}} Z_{2,j}$ for all $j \in [n]$, then the stochastic ordering is preserved under any multivariate nondecreasing function f, i.e., $f(Z_1) \leq f(Z_2)$ ([39], Example 9.2(a)). Since the functions $f(z) = z^r$ for nonnegative RV z and $f(z_1, \ldots, z_n) = \max(z_1, \ldots, z_n)$ are both nondecreasing, it follows that *i)* $\mathbb{E}[Z_1^r] \leq \mathbb{E}[Z_2^r]$, *ii)* $\mathbb{E}[\max(Z_{1,1}, \ldots, Z_{1,n})] \leq \mathbb{E}[\max(Z_{2,1}, \ldots, Z_{2,n})]$, and in fact $\mathbb{E}[\max(Z_{1,1}, \ldots, Z_{1,n})^r] \leq \mathbb{E}[\max(Z_{2,1}, \ldots, Z_{2,n})^r]$, for any nonnegative integer r.

Throughout the chapter we indicate the continuous RV matched to a discrete RV, say Z, by $\tilde{Z}$. As summarized in Table 3.4, let $\tilde{X} \sim \text{Exp}(1)$ denote a unit-rate exponential RV, and $\tilde{Y}^{(c)} \sim \text{Gamma}(c, 1) = \tilde{X}_1 + \cdots + \tilde{X}_c$ denote a Gamma RV, i.e., the sum of c independent unit-rate exponential RVs. For $\phi > 0$ it is easily seen that $\frac{1}{\phi}\tilde{X} \sim \text{Exp}(\phi)$ and $\frac{1}{\phi}\tilde{Y}^{(c)} \sim \text{Gamma}(c, \phi)$. Thus there is no loss in generality in restricting our attention to unit rate exponentials and Gamma RVs, as the general case is obtained by scaling.

Recall the discrete definitions of $X_{n:n} \equiv \max(X_1, \ldots, X_n)$ (for independent $X_j \sim \text{Geo}(q_j)$) and $Y_{n:n}^{(c)} \equiv \max(Y_1^{(c)}, \ldots, Y_n^{(c)})$ (for independent $Y_j^{(c)} = X_{j,1} + \cdots + X_{j,c}$ with $X_{j,k} \sim \text{Geo}(q_{j,k})$). Set $\phi(q) \equiv -\log(1-q)$. Define the continuous analog $\tilde{X}_{n:n} \equiv \max(\tilde{X}_1, \ldots, \tilde{X}_n)$ (for independent $\tilde{X}_j$), where *i)* under Assumption A2 $\tilde{X}_j \sim \frac{1}{\phi(q_j)}\tilde{X} \sim \text{Exp}(\phi(q_j))$, while *ii)* under Assumption A3 $\tilde{X}_j \sim \frac{1}{\phi(q)}\tilde{X} \sim \text{Exp}(\phi(q))$. Similarly, we define the continuous analog $\tilde{Y}_{n:n}^{(c)} \equiv \max(\tilde{Y}_1^{(c)}, \ldots, \tilde{Y}_n^{(c)})$ (for independent $\tilde{Y}_j^{(c)} = \tilde{X}_{j,1} + \cdots + \tilde{X}_{j,c}$ the sum of c independent RVs $(\tilde{X}_{j,k}, k \in [c])$). Now,

i) Under Assumption A1 $\tilde{X}_{j,k} \sim \frac{1}{\phi(q_{j,k})}\tilde{X} \sim \text{Exp}(\phi(q_{j,k}))$, and $\tilde{Y}_j^{(c)} \sim \text{HypoExp}(c, \phi(\mathbf{q}_j))$ for $\phi(\mathbf{q}_j) = (\phi(q_{j,k}), k \in [c])$;

ii) Under Assumption A2 $\tilde{X}_{j,k} \sim \frac{1}{\phi(q_j)}\tilde{X} \sim \text{Exp}(\phi(q_j))$, and $\tilde{Y}_j^{(c)} = \frac{1}{\phi(q_j)}\tilde{Y}^{(c)} \sim \text{Gamma}(c, \phi(q_j))$;

iii) Under Assumption A3 $\tilde{X}_{j,k} \sim \frac{1}{\phi(q)}\tilde{X} \sim \text{Exp}(\phi(q))$, and $\tilde{Y}_j^{(c)} = \frac{1}{\phi(q)}\tilde{Y}^{(c)} \sim \text{Gamma}(c, \phi(q))$.

We now establish several stochastic orderings between discrete RVs and their continuous analogs.

Proposition 1. *The following stochastic orderings hold for each $j \in [n]$:*

i) Under Assumption A1, for RLC: $\tilde{Y}_j^{(c)} \sim \text{HypoExp}(c, \phi(\mathbf{q}_j))$ *and* $Y_j^{(c)} \sim \text{GenNegBin}(c, \mathbf{q}_j)$*:*

$$\tilde{Y}_j^{(c)} \leq_{\text{st}} Y_j^{(c)} \leq_{\text{st}} \tilde{Y}_j^{(c)} + c. \tag{3.2}$$

ii) Under Assumption A2, for UT: $\tilde{X}_j \sim \frac{1}{\phi(q_j)}\tilde{X} \sim \text{Exp}(\phi(q_j))$ *and* $X_j \sim \text{Geo}(q_j)$*:*

$$\frac{1}{\phi(q_j)}\tilde{X} \leq_{\text{st}} X_j \leq_{\text{st}} \frac{1}{\phi(q_j)}\tilde{X} + 1. \tag{3.3}$$

iii) Under Assumption A2, for RLC: $\tilde{Y}_j^{(c)} \sim \frac{1}{\phi(q_j)}\tilde{Y}^{(c)} \sim \text{Gamma}(c, \phi(q_j))$ *and* $Y_j^{(c)} \sim \text{NegBin}(c, q_j)$*:*

$$\frac{1}{\phi(q_j)}\tilde{Y}^{(c)} \leq_{\text{st}} Y_j^{(c)} \leq_{\text{st}} \frac{1}{\phi(q_j)}\tilde{Y}^{(c)} + c. \tag{3.4}$$

iv) Under Assumption A3, for UT: $\tilde{X}_j \sim \frac{1}{\phi(q)}\tilde{X} \sim \text{Exp}(\phi(q))$ *and* $X \sim \text{Geo}(q)$*:*

$$\frac{1}{\phi(q)}\tilde{X} \leq_{\text{st}} X \leq_{\text{st}} \frac{1}{\phi(q)}\tilde{X} + 1. \tag{3.5}$$

v) Under Assumption A3, for RLC: $\tilde{Y}_j^{(c)} \sim \frac{1}{\phi(q)}\tilde{Y}^{(c)} \sim \text{Gamma}(c, \phi(q))$ *and* $Y^{(c)} \sim \text{NegBin}(c, q)$*:*

$$\frac{1}{\phi(q)}\tilde{Y}^{(c)} \leq_{\text{st}} Y^{(c)} \leq_{\text{st}} \frac{1}{\phi(q)}\tilde{Y}^{(c)} + c. \tag{3.6}$$

The above stochastic orderings omit the dependence upon j *when the distributions are not dependent upon* j*. Furthermore, all these stochastic orderings are preserved when the RVs are raised to the* r^{th} *power, and, after taking expectations of these powers, the ordering is preserved for the* r^{th} *moments.*

Proof. It suffices to prove cases *i*) and *ii*), since *iii*) and *v*) are special cases of *i*), and *iv*) is a special case of *ii*). We first prove case *ii*). Let $X_j \sim \text{Geo}(q_j)$ and recall $\tilde{X} \sim \text{Exp}(1)$. We first show $\frac{1}{\phi(q_j)}\tilde{X} \leq_{\text{st}} X$. Observe for any $x \geq 0$:

$$\bar{F}_{\frac{1}{\phi(q_j)}\tilde{X}}(x) = \mathbb{P}\left(\tilde{X} > \phi(q_j)x\right) = \mathrm{e}^{-\phi(q_j)x} = (1-q_j)^x \leq (1-q_j)^{\lfloor x \rfloor} = \mathbb{P}(X_j > x) = \bar{F}_{X_j}(x), \tag{3.7}$$

where $\lfloor x \rfloor$ is the largest integer not exceeding x. We next show $X_j \leq_{\text{st}} \frac{1}{\phi(q_j)}\tilde{X} + 1$:

$$\bar{F}_{X_j}(x) = \mathbb{P}(X_j > x) = (1-q_j)^{\lfloor x \rfloor} < (1-q_j)^{(x-1)} = \mathrm{e}^{-\phi(q_j)(x-1)} = \mathbb{P}(\tilde{X} > \phi(q_j)(x-1)) = \bar{F}_{\frac{1}{\phi(q_j)}\tilde{X}+1}(x). \tag{3.8}$$

We next prove case i). The fact $\tilde{Y}_j^{(c)} \leq_{\text{st}} Y_j^{(c)} \leq_{\text{st}} \tilde{Y}_j^{(c)} + c$ follows from 1) the stochastic ordering of $\frac{1}{\phi(q_{j,k})}\tilde{X} \leq_{\text{st}} X_{j,k} \leq_{\text{st}} \frac{1}{\phi(q_{j,k})}\tilde{X} + 1$, 2) the fact that the sum of nonnegative RVs is a nondecreasing function of a random vector composed of those random variables, and 3) the fact that stochastic ordering is preserved under nondecreasing functions of random vectors with independent stochastically ordered components. □

3.3 Delay under uncoded transmission

The r^{th} ($r \in \mathbb{N}$) moment of delay under UT, $\mathbb{E}[X_{n:n}^r]$, is investigated in this section. We first give lower and upper bounds (§3.3.1), then derive an exact closed-form expression using (moment) generating function (§3.3.2). Finally this section closes with further remarks on related work. Throughout this section we assume A2: state-independent receptions, heterogeneous receivers. State-independence follows from the fact that under UT there is no coding.

3.3.1 Lower and upper bounds on moments of the delay under UT

Prop. 3 below hinges upon the stochastic ordering between exponential and geometric RVs, and the following "min-max identity". Recall $[n]_s$ is the set of all subsets of $[n]$ of size s.

Proposition 2 (min-max identity,[2], pp. 129)**.** *For nonnegative (not necessarily independent) RVs $Z_1, \ldots, Z_n$:*

$$\mathbb{E}\left[\max_{j\in[n]} Z_j\right] = \sum_{s=1}^{n} (-1)^{s+1} \sum_{A\in[n]_s} \mathbb{E}\left[\min_{j\in A} Z_j\right] \tag{3.9}$$

Proposition 3. *Assume A2: State-independent receptions, heterogeneous receivers. The r^{th} moment of the maximum of n independent geometric RVs $\mathbb{E}[X_{n:n}^r]$ has lower and upper bounds*

$$\psi_r(\mathbf{q}) \leq \mathbb{E}[X_{n:n}^r] \leq \sum_{s=0}^{r} \binom{r}{s} \psi_s(\mathbf{q}), \tag{3.10}$$

where

$$\psi_r(\mathbf{q}) \equiv r! \sum_{s=1}^{n} (-1)^{s+1} \sum_{A \in [n]_s} \left(\sum_{j \in A} \phi(q_j) \right)^{-r}. \tag{3.11}$$

Proof. It follows from the stochastic ordering presented in §3.2 that

$$\mathbb{E}[\tilde{X}_{n:n}^r] \leq \mathbb{E}[X_{n:n}^r] \leq \mathbb{E}[(\tilde{X}_{n:n} + 1)^r]. \tag{3.12}$$

Applying the min-max identity (Prop. 2) to $(\tilde{X}_1^r, \ldots, \tilde{X}_n^r)$ yields

$$\mathbb{E}[\max_{j \in [n]} \tilde{X}_j^r] = \sum_{s=1}^{n} (-1)^{s+1} \sum_{A \in [n]_s} \mathbb{E}[\min_{j \in A} \tilde{X}_j^r]. \tag{3.13}$$

Since *i*) the minimum of independent exponential RVs is an exponential RV whose rate is the sum of the rates, and *ii*) the r^{th} moment of an exponential RV with rate λ (say) is $r!/\lambda^r$, it follows that $\mathbb{E}[\min_{j \in A} \tilde{X}_j^r] = r! \left(\sum_{j \in A} \phi(q_j) \right)^{-r}$. This gives the lower bound $\psi_r(\mathbf{q})$ in (3.10). The binomial theorem gives

$$\left(\max_{j \in [n]} \tilde{X}_j + 1 \right)^r = \sum_{s=0}^{r} \binom{r}{s} \max_{j \in [n]} \tilde{X}_j^s, \tag{3.14}$$

and taking expectations gives the upper bound in (3.10). □

Remark 8. *Notice the gap is given by:* $\sum_{s=0}^{r-1} \binom{r}{s} \psi_s(\mathbf{q})$*, and when* $r = 1$ *the gap equals* 1*, since* $\psi_0(\mathbf{q}) = \sum_{s=1}^{n} (-1)^{s+1} \binom{n}{s} = -\left[\sum_{s=0}^{n} (-1)^s \binom{n}{s} - 1 \right] = 1$.

Remark 9. *Lemma 4 of*[33] *derives a lower bound on the expected time slots required to broadcast a packet to all* n *homogeneous receivers (i.e.,* $\mathbb{E}[X_{n:n}]$ *when* $q_j = q$ *for* $j \in [n]$*), in order to construct an outer bound on the stability region of scheduling policies (Theorem 6). Our result is more general in that it i) provides upper and lower bounds, ii) applies to heterogeneous channels* $\mathbf{q} = (q_1, \ldots, q_n)$*, and iii) yields arbitrary (*r^{th}*) moments. Further, the proof technique we employ is probabilistic instead of analytic.*

Remark 10. *The classic (sequential) coupon-collector problem (*[40] *§3.6) asks for the expected time to receive all of* n *coupons, when in each time slot a new coupon is selected uniformly at random*

from $[n]$*. The broadcast delay problem can be considered as a "parallel" coupon-collector problem in that at each time slot each of the* n *coupons is received independently with probabilities* $\mathbf{q}$*, i.e., a reception by receiver* j *corresponds to collecting a coupon of type* j*.*

Specializing Prop. 3 to the homogeneous channel case with $r = 1, q_j = q$ for all $j \in [n]$ yields simpler expressions for the lower and upper bounds after using the combinatorial identity $\sum_{s=1}^{n}(-1)^s\binom{n}{s}(-\frac{1}{s}) = \sum_{j=1}^{n}\frac{1}{j}$. We observe the lower bound in Cor. 1 is the same as the one from Lemma 4 in[33].

Corollary 1. *Assume A3: State-independent receptions, homogeneous receivers. The expectation of the maximum of* n *iid geometric RVs with parameter* q *is bounded as:*

$$\frac{H_n}{\phi(q)} \leq \mathbb{E}[X_{n:n}] \leq \frac{H_n}{\phi(q)} + 1.$$

for $H_n \equiv 1 + 1/2 + \cdots + 1/n$ *the* n^{th} *harmonic number.*

3.3.2 An exact moment expression for the delay under UT

In this subsection we obtain an exact moment expression via the min-max identity (Prop. 2) and the property that the minimum of independent geometric RVs is also a geometric RV.

Proposition 4. *Assume A2: State-independent receptions, heterogeneous receivers. The* r^{th} *moment of the maximum of* n *independent geometric RVs* $X_j \sim \text{Geo}(q_j), j \in [n]$ *is given by:*

$$\mathbb{E}[X_{n:n}^r] = \sum_{s=1}^{n}(-1)^{s+1}\sum_{A\in[n]_s}\frac{1}{q_A}\sum_{l=1}^{r}\left\{{r \atop l}\right\}\left(\frac{1}{q_A}-1\right)^{l-1} l!, \tag{3.15}$$

where $q_A \equiv 1 - \prod_{j\in A}(1-q_j)$ *and* $\left\{{r \atop l}\right\}$ *denotes the Stirling number of the second kind.*

Proof. Use min-max identity:

$$\begin{aligned}
\mathbb{E}[X_{n:n}^r] &= \mathbb{E}[(\max(X_1,\ldots,X_n))^r] \\
&= \mathbb{E}[\max(X_1^r,\ldots,X_n^r)] \\
&= \sum_{s=1}^{n}(-1)^{s+1}\sum_{A\in[n]_s}\mathbb{E}\left[\min_{j\in A}X_j^r\right] \\
&= \sum_{s=1}^{n}(-1)^{s+1}\sum_{A\in[n]_s}\mathbb{E}\left[\left(\min_{j\in A}X_j\right)^r\right] \qquad (3.16)
\end{aligned}$$

Now recognize $\min_{j\in A} X_j \sim \text{Geo}(q_A)$ (due to independence) and substitute the expression for the r^{th} moment of Geo(q) derived in Prop. 5 shown below. □

The following corollary illustrates the expressions from Prop. 4 for the first two moments ($r = 1, 2$).

Corollary 2. *Assume A2: State-independent receptions, heterogeneous receivers. The first two moments of maximum of n independent geometric RVs with parameters* $\mathbf{q}$ *are:*

$$\begin{aligned}
\mathbb{E}[X_{n:n}] &= \sum_{s=1}^{n}(-1)^{s+1}\sum_{A\in[n]_s}\left(1-\prod_{j\in A}(1-q_j)\right)^{-1} \\
\mathbb{E}[X_{n:n}^2] &= \sum_{s=1}^{n}(-1)^{s+1}\sum_{A\in[n]_s}\frac{1+\prod_{j\in A}(1-q_j)}{\left(1-\prod_{j\in A}(1-q_j)\right)^2}.
\end{aligned}$$

The following proposition gives the r^{th} moment of a geometric RV.

Proposition 5. *The* r^{th} *($r \in \mathbb{N}$) moment of a geometric RV* $X \sim \text{Geo}(q)$ *is*

$$\mathbb{E}[X^r] = \frac{1}{q}\sum_{l=1}^{r}\left\{{r \atop l}\right\}\left(\frac{1}{q}-1\right)^{l-1} l!,\ q\in(0,1). \qquad (3.17)$$

where $\left\{{r \atop l}\right\}$ *is the Stirling number of the second kind.*

Proof. Let $M_X(t) = \frac{qe^t}{1-(1-q)e^t}$ for $t < -\log(1-q)$ denote the moment generating function (MGF)

for the geometric distribution. We first establish the r^{th} derivative of the MGF is given by

$$M_X^{(r)}(t) = \sum_{l=1}^{r} a_l^{(r)} \left(\frac{\mathrm{d}^l}{\mathrm{d}s^l} \tilde{M}_X(s) \right) (s-1)^l \Bigg|_{s=s(t)} \tag{3.18}$$

where $s(t) \equiv 1-(1-q)\mathrm{e}^t$, and $\tilde{M}_X(s) \equiv \frac{q}{1-q}\left(s^{-1}-1\right)$. It follows that $\tilde{M}_X(s(t)) = M_X(t)$, and $\frac{\mathrm{d}s(t)}{\mathrm{d}t} = s(t)-1$, and that the coefficients $a_l^{(r)}$ satisfy the following recurrence:

$$a_l^{(r)} = l \cdot a_l^{(r-1)} + a_{l-1}^{(r-1)},\ l \in [r], \tag{3.19}$$

with boundary conditions $a_l^{(r)} = 0$ if $l > r$, and $a_0^{(r)} = \mathbb{I}_{\{r=0\}}$ (here $\mathbb{I}$ is the indicator function).

Observe $a_r^{(r)} = 1 = a_1^{(r)}$. We prove by induction in r. The base case $r = 1, 2$ can be verified to be true, with $M_X^{(1)}(t) = \frac{q}{1-q}\left(-s(t)^{-1}+s(t)^{-2}\right), M_X^{(2)}(t) = \frac{q}{1-q}\left(s(t)^{-1}-3s(t)^{-2}+2s(t)^{-3}\right)$. Assuming (3.18) holds for r, the correctness of (3.18) for $r+1$ can be seen by applying the chain rule: $\frac{\mathrm{d}}{\mathrm{d}t} M_X^{(r)}(t) = \left(\frac{\mathrm{d}}{\mathrm{d}s} M_X^{(r)}(t)\right) \cdot \frac{\mathrm{d}s}{\mathrm{d}t}$.

Now notice the recurrence of $a_l^{(r)}$ (together with the boundary conditions) exactly coincides with that of the Stirling number of the second kind for which this recurrence can be solved using generating functions ([20], §1.6):

$$a_l^{(r)} = \left\{ {r \atop l} \right\} = \sum_{m=1}^{l} (-1)^{l-m} \frac{m^r}{m!\,(l-m)!},\ r, l \geq 0. \tag{3.20}$$

Observe that $\frac{\mathrm{d}^l}{\mathrm{d}s^l} \tilde{M}_X(s) = \frac{q}{1-q}(-1)^l\, l! \cdot s^{-(l+1)}$. Substitution and evaluation at $t = 0$ yields the desired formula. □

We essentially apply the chain rule to higher-order derivatives. With an appropriate change of variable, the r^{th} derivative of the original MGF (in terms of t) is expressed as a weighted sum of all lower-order derivatives with respect to the new variable s, where these lower-order derivatives are easy to characterize and the coefficients obey a recurrence which can be further solved.

It is interesting to observe Stirling numbers of the second kind (often defined as the number of partitions of $[r]$ into l sets) often show up in expressions for moments of discrete RVs. As

another example, the r^{th} moment of a Poisson RV with parameter λ, say $Z \sim \text{Po}(\lambda)$ is given by $\mathbb{E}[Z^r] = \sum_{l=1}^{r} \left\{ {r \atop l} \right\} \lambda^l$. Interpretations of the Stirling numbers of the second kind including their affinity to differential operators are discussed in[41].

3.3.3 Further remarks on related work

Prop. 4 builds upon the min-max identity in Prop. 2 from[2]. It is straightforward to use the principle of inclusion and exclusion (PIE) upon which Prop. 2 is proved to establish a sequence of lower and upper bounds on $\mathbb{E}[Z_{n:n}]$. Moreover, there is an analogous "max-min" identity giving $\mathbb{E}[\min_{j\in[n]} Z_j]$ in terms of $\mathbb{E}[\max_{j\in A} Z_j]$ for $A \subseteq [n]$. Finally, recent work[42] has generalized the min-max and max-min identities to a more general "sorting" identity in a non-probabilistic setting.

Characterizing the expectation of the maximum of n geometric RVs is addressed in[43]. Let $(X_1, \ldots, X_n)$ be iid geometric RVs with parameter q. By using Fourier analysis of the distribution of the fractional part of the maximum of corresponding iid exponential RVs,[43] (Corollary 2) obtains an exact formula:

$$\mathbb{E}[X_{n:n}] = \frac{1}{2} + \frac{H_n}{\phi(q)} - \sum_{k\neq 0} \frac{1}{2\pi k \mathrm{i}} \prod_{j=1}^{n} \left(1 + \frac{2\pi k \mathrm{i}}{j\phi(q)}\right)^{-1}, \tag{3.21}$$

where $\mathrm{i} = \sqrt{-1}$. This result improves an earlier result[44] (Eq. 2.8). Using Corollary 2, the infinite sum in (3.21) can be shown to have absolute value no larger than 1/2.

3.4 Delay under random linear combinations

We now turn our attention from delay under UT to delay under RLC, with a focus on the r^{th} ($r \in \mathbb{N}$) moment, $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$. We provide *i*) lower and upper bounds for $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ in §3.4.1 (Props. 6, 7, and 8), *ii*) exact expressions for $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ in §3.4.2 (Prop. 9) which involves a sum over an infinite number of terms, *iii*) a recurrence for $Y_{n:n}$ in §3.4.3 (Prop. 10) which allows $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ to be computed in a finite number of steps, and *iv*) in §3.4.4 a characterization of the channel reception probabilities $\mathbf{q} = (q_1, \ldots, q_n)$ for which $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ is minimized.

3.4.1 Lower and upper bounds on moments of the delay under RLC

Our first result, Prop. 6, holds for the most general case (Assumption A1), and follows immediately from the stochastic ordering between generalized negative binomial (sum of independent geometric) and hypo-exponential (sum of independent exponential) RVs. The drawback to this result is that the upper bound is loose, i.e., $Y_j^{(c)} \leq_{\text{st}} \tilde{Y}_j^{(c)} + c$. To address this, we present in Prop. 7 a continuous RV $\tilde{W}^{(c)}$ such that $\tilde{W}^{(c)} \leq_{\text{st}} Y^{(c)} \leq_{\text{st}} \tilde{W}^{(c)} + 1$, with the caveat that the RV $\tilde{W}^{(c)}$ is constructed only under Assumption A2, in which case the hypo-exponential and the generalized negative binomial RVs reduces to gamma and negative binomial RVs respectively. Finally, we present in Prop. 8 explicit lower and upper bounds on delay moments under Assumption A3.

Proposition 6. *Assume A1: State-dependent receptions, heterogeneous receivers. The* r^{th} *moment of the maximum of* n *independent generalized negative binomial RVs* $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ *has lower and upper bounds*

$$\Psi_r(\mathbf{Q}) \leq \mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] \leq \sum_{s=0}^{r} \binom{r}{s} \Psi_s(\mathbf{Q}) c^{r-s}, \tag{3.22}$$

where

$$\Psi_r(\mathbf{Q}) \equiv \mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right] = \int_0^\infty \left(1 - \prod_{j\in[n]} F_{\tilde{Y}_j^{(c)}}\left(y^{\frac{1}{r}}\right)\right) \mathrm{d}y, \tag{3.23}$$

and $F_{\tilde{Y}_j^{(c)}}(y)$ *is the CDF for the hypo-exponential RV* $\tilde{Y}_j^{(c)} \sim \text{HypoExp}(c, \phi(\mathbf{q}_j))$*, with parameter vector* $\phi(\mathbf{q}_j) = (\phi(q_{j,1}), \ldots, \phi(q_{j,c}))$.

Proof. Recall the stochastic ordering discussed in §3.2: $\left(\tilde{Y}_{n:n}^{(c)}\right)^r \leq_{\text{st}} \left(Y_{n:n}^{(c)}\right)^r \leq_{\text{st}} \left(\tilde{Y}_{n:n}^{(c)} + c\right)^r$. Taking expectations of these stochastically ordered RVs (and using the binomial theorem on the upper

bound) yields (3.22). To obtain (3.23) observe

$$\begin{aligned}
\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right] &= \int_0^\infty \mathbb{P}\left(\left(\tilde{Y}_{n:n}^{(c)}\right)^r > y\right) \mathrm{d}y \\
&= \int_0^\infty \left(1 - \mathbb{P}\left(\left(\tilde{Y}_{n:n}^{(c)}\right)^r \leq y\right)\right) \mathrm{d}y \\
&= \int_0^\infty \left(1 - \prod_{j\in[n]} \mathbb{P}\left(\left(\tilde{Y}_{j}^{(c)}\right)^r \leq y\right)\right) \mathrm{d}y \\
&= \int_0^\infty \left(1 - \prod_{j\in[n]} \mathbb{P}\left(\tilde{Y}_{j}^{(c)} \leq y^{\frac{1}{r}}\right)\right) \mathrm{d}y. \qquad (3.24)
\end{aligned}$$

□

Proposition 7. *Assume A2: State-independent receptions, heterogeneous receivers. The r^{th} moment of the maximum of n independent generalized negative binomial RVs $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ has lower and upper bounds*

$$\tilde{\Psi}_r(\mathbf{q}) \leq \mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] \leq \sum_{s=0}^{r} \binom{r}{s} \tilde{\Psi}_s(\mathbf{q}), \qquad (3.25)$$

where

$$\tilde{\Psi}_r(\mathbf{q}) \equiv \mathbb{E}\left[\left(\tilde{W}_{n:n}^{(c)}\right)^r\right] = c^r + \int_{c^r}^\infty \left(1 - \prod_{j\in[n]} I_{q_j}(c, w^{\frac{1}{r}} - c + 1)\right) \mathrm{d}w. \qquad (3.26)$$

Proof. For integer blocklength c and success probability $q \in (0,1)$, define the continuous RV $\tilde{W}^{(c)}$ with support $[c,\infty)$ and CDF $F_{\tilde{W}^{(c)}}(w) = I_q(c, w-c+1)\mathbf{1}_{\{w\geq c\}}$. We first show that $\tilde{W}^{(c)}$ is a valid continuous RV. Clearly $F_{\tilde{W}^{(c)}}(w) \in [0,1]$, and by construction $\lim_{w\to-\infty} F_{\tilde{W}^{(c)}}(w) = 0$. By Lem. 3 proved in the appendix we have: $\frac{\mathrm{d}}{\mathrm{d}w} F_{\tilde{W}^{(c)}}(w) > 0$, and it also follows that $\lim_{w\to\infty} F_{\tilde{W}^{(c)}}(w) = 1$.

We now show that $\tilde{W}^{(c)} \leq_{\text{st}} Y^{(c)} \leq_{\text{st}} \tilde{W}^{(c)} + 1$ for $Y^{(c)} \sim \text{NegBin}(c,q)$. Observe that the CDF for $\tilde{W}^{(c)}$ and $Y^{(c)}$ are equal at every integer $m \geq c$:

$$F_{\tilde{W}^{(c)}}(m) = I_q(c, m-c+1) = \mathbb{P}(\text{Bin}(m,q) \geq c) = \mathbb{P}(Y^{(c)} \leq m) = F_{Y^{(c)}}(m). \qquad (3.27)$$

This, along with the facts that $\frac{\mathrm{d}}{\mathrm{d}w} F_{\tilde{W}^{(c)}}(w) > 0$ and $F_{Y^{(c)}}(w)$ is piecewise constant in w guarantee

$F_{\tilde{W}^{(c)}}(w) \geq F_{Y^{(c)}}(w)$, and thus $\tilde{W}^{(c)} \leq_{\text{st}} Y^{(c)}$. Similarly,

$$F_{\tilde{W}^{(c)}+1}(m) = I_q(c,(m-1)-c+1) = \mathbb{P}(\text{Bin}(m-1,q) \geq c) = \mathbb{P}(Y^{(c)} \leq m-1) = F_{Y^{(c)}}(m-1), \quad (3.28)$$

and thus $F_{\tilde{W}^{(c)}+1}(m) = F_{Y^{(c)}}(m-1) \leq F_{Y^{(c)}}(m)$, which establishes $Y^{(c)} \leq_{\text{st}} \tilde{W}^{(c)} + 1$. It follows that $\left(\tilde{W}^{(c)}_{n:n}\right)^r \leq_{\text{st}} \left(Y^{(c)}_{n:n}\right)^r \leq_{\text{st}} \left(\tilde{W}^{(c)}_{n:n}+1\right)^r$. Taking expectations of these stochastically ordered RVs (and using the binomial theorem on the upper bound) yields (3.25). To obtain (3.26) observe

$$\mathbb{E}\left[\left(\tilde{W}^{(c)}_{n:n}\right)^r\right] = \int_0^\infty \mathbb{P}\left(\left(\tilde{W}^{(c)}_{n:n}\right)^r > w\right) \mathrm{d}w = c^r + \int_{c^r}^\infty \mathbb{P}\left(\left(\tilde{W}^{(c)}_{n:n}\right)^r > w\right) \mathrm{d}w, \quad (3.29)$$

and apply the same steps used at the end of the proof of Prop. 6. □

Observe the bounds given in the previous two propositions may not be easy to compute since they themselves involve calculation of moments of the maximum order statistic of a continuous RV, the support of which may be infinite. The following proposition shows that by further bounding the continuous RV we obtain bounds that also have simple structure. More precisely, we will be working with the Gamma distribution, for which the bound can be made asymptotically tight in c when n is fixed (§3.5.3), and in n when c is fixed (§3.6). For simplicity, the following proposition is stated under Assumption A3, yet generalizing to heterogeneous receivers case is not hard.

Proposition 8. *Assume A3: State-independent receptions, homogeneous receivers. The* r^{th} *moment of the maximum of* n *iid negative binomial RVs* $\mathbb{E}\left[\left(Y^{(c)}_{n:n}\right)^r\right]$ *has lower and upper bounds*

$$\frac{1}{\phi(q)^r} l_{\tilde{Y}}(t,n,c,r) \leq \mathbb{E}\left[\left(Y^{(c)}_{n:n}\right)^r\right] \leq \sum_{p=0}^{r} \binom{r}{p} \frac{1}{\phi(q)^p} u_{\tilde{Y}}(s,n,c,p) c^{r-p}, \quad (3.30)$$

where

$$\begin{aligned} l_{\tilde{Y}}(t,n,c,r) &\equiv t - \left(t - \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]\right)\left(1 - Q(c, t^{\frac{1}{r}})\right)^{n-1} \\ u_{\tilde{Y}}(s,n,c,r) &\equiv s + n\left(\left(\frac{\Gamma(c+r)}{\Gamma(c)} - s\right) Q(c, s^{\frac{1}{r}}) + \frac{\Gamma(c+r)}{\Gamma(c)} s^{\frac{c+r-1}{r}} \mathrm{e}^{-s^{\frac{1}{r}}} \sum_{m=0}^{r-1} \frac{1}{s^{\frac{m}{r}}\Gamma(c+r-m)}\right) \end{aligned} \quad (3.31)$$

For the lower bound, $t \geq \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]$ *and* $\tilde{Y}^{(c)} \sim \text{Gamma}(c,1)$. *For the upper bound,* $s \geq 0$, *and the optimal* s^* *is such that* $nQ\left(c, (s^*)^{\frac{1}{r}}\right) = 1$ *for* $Q(c,s) \equiv \frac{\Gamma(c,s)}{\Gamma(c)} = \mathbb{P}\left(\tilde{Y}^{(c)} > s\right)$ *the CCDF of the* $\text{Gamma}(c,1)$ *distribution. For* $r = 1$, *the optimal upper bound is*

$$\mathbb{E}\left[Y_{n:n}^{(c)}\right] \leq c\left(1 + nQ\left(c+1, Q^{-1}\left(c, \frac{1}{n}\right)\right)\right). \tag{3.32}$$

Proof. The ordering $\frac{1}{\phi(q)^r}\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right] \leq \mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] \leq \sum_{p=0}^{r}\binom{r}{p}\frac{1}{\phi(q)^p}\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^p\right]c^{r-p}$ follows from Prop. 1. It remains to show the bounds as given in (3.31) are valid. The bounds we employ are introduced in Appendix §3.7.6. The lower bound is a direct application of (3.171) to $\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]$ by further recognizing $\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r > t\right) = \mathbb{P}\left(\tilde{Y}^{(c)} > t^{\frac{1}{r}}\right) = Q(c, t^{\frac{1}{r}})$. The upper bound is a direct application of (3.172) to $\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]$, followed by an exchange of the order of integration, and repeated use of the recursion $Q(c+1,x) = Q(c,x) + x^c\mathrm{e}^{-x}/\Gamma(c+1)$. In particular, applying the upper bound and exchanging the order of integration gives:

$$\begin{aligned}
\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right] &\leq s + n\int_s^\infty \mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r > t\right)\mathrm{d}t \\
&= s + n\int_s^\infty Q(c, t^{\frac{1}{r}})\mathrm{d}t \\
&= s + \frac{n}{\Gamma(c)}\int_s^\infty\left(\int_{t^{\frac{1}{r}}}^\infty u^{c-1}\mathrm{e}^{-u}\mathrm{d}u\right)\mathrm{d}t \\
&= s + \frac{n}{\Gamma(c)}\int_{s^{\frac{1}{r}}}^\infty\left(\int_s^{u^r}\mathrm{d}t\right)u^{c-1}\mathrm{e}^{-u}\mathrm{d}u \\
&= s + \frac{n}{\Gamma(c)}\int_{s^{\frac{1}{r}}}^\infty (u^r - s)u^{c-1}\mathrm{e}^{-u}\mathrm{d}u
\end{aligned} \tag{3.33}$$

Now application of the recursion mentioned above gives

$$\begin{aligned}
\mathbb{E}\left[\left(\tilde{Y}^{(c)}_{n:n}\right)^r\right] &\leq s + \frac{n}{\Gamma(c)}\left(\int_{s^{\frac{1}{r}}}^{\infty} u^{r+c-1}\mathrm{e}^{-u}\mathrm{d}u - s\int_{s^{\frac{1}{r}}}^{\infty} u^{c-1}\mathrm{e}^{-u}\mathrm{d}u\right) \\
&= s + \frac{n}{\Gamma(c)}\left(\Gamma(c+r, s^{\frac{1}{r}}) - s\Gamma(c, s^{\frac{1}{r}})\right) \\
&= s + n\left(\frac{\Gamma(c+r)}{\Gamma(c)}Q(c+r, s^{\frac{1}{r}}) - sQ(c, s^{\frac{1}{r}})\right) \\
&= s + n\left(\frac{\Gamma(c+r)}{\Gamma(c)}\left(Q(c, s^{\frac{1}{r}}) + \left(s^{\frac{1}{r}}\right)^{c+r-1}\mathrm{e}^{-s^{\frac{1}{r}}}\sum_{m=0}^{r-1}\frac{1}{\left(s^{\frac{1}{r}}\right)^m\Gamma(c+r-m)}\right) - sQ(c, s^{\frac{1}{r}})\right) \\
&= s + n\left(\left(\frac{\Gamma(c+r)}{\Gamma(c)} - s\right)Q(c, s^{\frac{1}{r}}) + \frac{\Gamma(c+r)}{\Gamma(c)}s^{\frac{c+r-1}{r}}\mathrm{e}^{-s^{\frac{1}{r}}}\sum_{m=0}^{r-1}\frac{1}{s^{\frac{m}{r}}\Gamma(c+r-m)}\right) \qquad (3.34)
\end{aligned}$$

The optimal s^* may be found as:

$$n\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r > s^*\right) = 1 \;\Rightarrow\; nQ\left(c, (s^*)^{\frac{1}{r}}\right) = 1 \;\Rightarrow\; s^* = \left(Q^{-1}\left(c, \frac{1}{n}\right)\right)^r. \qquad (3.35)$$

Here, $Q^{-1}(c, u) = x$ is the inverse with respect to x of $Q(c, x)$ so that $Q(c, x) = u$, for $u \in [0, 1]$. Substitution of the optimal s^* into the above bound gives

$$\begin{aligned}
&s^* + n\left(\left(\frac{\Gamma(c+r)}{\Gamma(c)} - s^*\right)Q(c, (s^*)^{\frac{1}{r}}) + \frac{\Gamma(c+r)}{\Gamma(c)}\left((s^*)^{\frac{1}{r}}\right)^{c+r-1}\mathrm{e}^{-(s^*)^{\frac{1}{r}}}\sum_{m=0}^{r-1}\frac{1}{\left((s^*)^{\frac{1}{r}}\right)^m\Gamma(c+r-m)}\right) \\
&= \left(Q^{-1}\left(c, \frac{1}{n}\right)\right)^r + \frac{\Gamma(c+r)}{\Gamma(c)} - \left(Q^{-1}\left(c, \frac{1}{n}\right)\right)^r + n\frac{\Gamma(c+r)}{\Gamma(c)}\left((s^*)^{\frac{1}{r}}\right)^{c+r-1}\mathrm{e}^{-(s^*)^{\frac{1}{r}}}\sum_{m=0}^{r-1}\frac{1}{\left((s^*)^{\frac{1}{r}}\right)^m\Gamma(c+r-m)} \\
&= \frac{\Gamma(c+r)}{\Gamma(c)} + n\frac{\Gamma(c+r)}{\Gamma(c)}(s^*)^{\frac{c+r-1}{r}}\mathrm{e}^{-(s^*)^{\frac{1}{r}}}\sum_{m=0}^{r-1}\frac{1}{(s^*)^{\frac{m}{r}}\Gamma(c+r-m)}. \qquad (3.36)
\end{aligned}$$

For $r = 1$, the above expression simplifies to $c + nc\frac{(s^*)^c\mathrm{e}^{-s^*}}{\Gamma(c+1)}$, which after applying the recursion can be written as (3.32). □

Observe we have used two distinct continuous-RV stochastic orderings in the last two propositions. First, Prop. 7 relied upon $\tilde{W}^{(c)}_j \leq_{\mathrm{st}} Y^{(c)}_j \leq_{\mathrm{st}} \tilde{W}^{(c)}_j + 1$ under Assumption A2, while Prop. 8 relied upon $\tilde{Y}^{(c)}_j \leq_{\mathrm{st}} Y^{(c)}_j \leq_{\mathrm{st}} \tilde{Y}^{(c)}_j + c$ under Assumption A3. These orderings are illustrated in Fig. 3.2. Note the stochastic ordering with $\tilde{W}^{(c)}_j$ is much tighter to $Y^{(c)}_j$ than is that with $\tilde{Y}^{(c)}_j$, although

the latter is much easier to compute than the former.

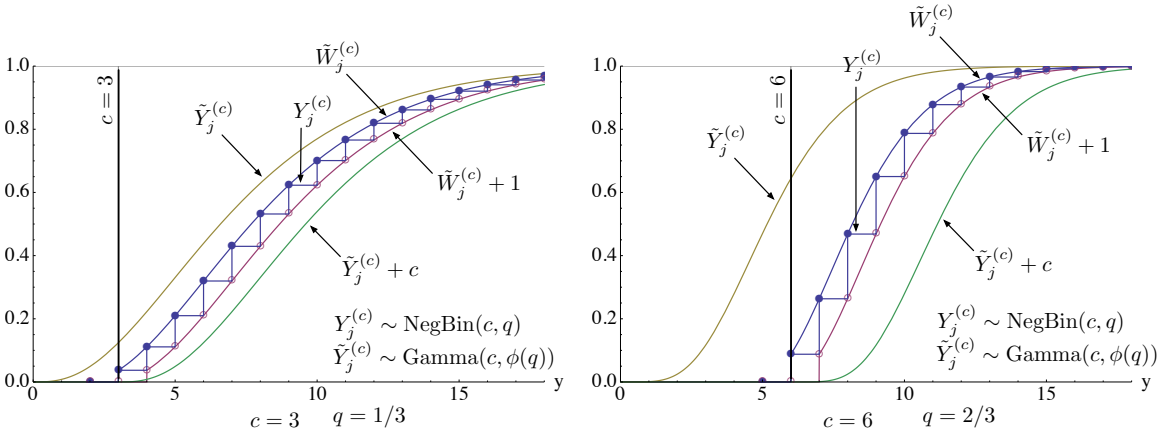

Figure 3.2: Illustration of the CDFs for RVs in the stochastic orderings $\tilde{W}_j^{(c)} \leq_{\text{st}} Y_j^{(c)} \leq_{\text{st}} \tilde{W}_j^{(c)} + 1$ and $\tilde{Y}_j^{(c)} \leq_{\text{st}} Y_j^{(c)} \leq_{\text{st}} \tilde{Y}_j^{(c)} + c$ for $c = 3$ and $q = 1/3$ (left), and $c = 6$ and $q = 2/3$ (right).

3.4.2 Exact expressions of the moments of delay under RLC

Prop. 9 gives an expression for the exact delay $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ under RLC for the state-dependent case as well as two equivalent expressions for $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ under RLC for the state-independent case.

Proposition 9. *Assume A1: State-dependent receptions, heterogeneous receivers:*

$$\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] = c^r + \sum_{m=c^r}^{\infty} \left(1 - \prod_{j=1}^{n} \sum_{t=c}^{\lfloor m^{\frac{1}{r}} \rfloor} \sum_{\boldsymbol{\alpha} \in \mathcal{A}_t} \prod_{k=1}^{c} (1 - q_{jk})^{\alpha_k - 1} q_{jk}\right), \tag{3.37}$$

where $\mathcal{A}_t$ is the finite set of all c-vectors of positive integers that sum to t, i.e.,

$$\mathcal{A}_t = \left\{\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_c) \ : \ \alpha_k \in \mathbb{N}, \ \sum_{k=1}^{c} \alpha_k = t\right\}. \tag{3.38}$$

Assume A2: State-independent receptions, heterogeneous receivers. Eq. (3.37) *simplifies to*

$$\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] = c^r + \sum_{m=c^r}^{\infty} \left(1 - \prod_{j=1}^{n} \sum_{t=c}^{\lfloor m^{\frac{1}{r}} \rfloor} \binom{t-1}{c-1} (1 - q_j)^{t-c} q_j^c\right). \tag{3.39}$$

An alternative expression for the state-independent case is

$$\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] = c^r + \sum_{m=c^r}^{\infty}\left(1 - \prod_{j=1}^{n} I_{q_j}\left(c, \lfloor m^{\frac{1}{r}}\rfloor - c + 1\right)\right), \tag{3.40}$$

where in the last equation, $I_x(\alpha,\beta)$ is the regularized incomplete Beta function, which can be used as both the CDF of the Beta distribution and the tail probability of the binomial RV $W \sim \mathrm{Bin}(m,q)$:

$$\mathbb{P}(W \geq l) = \sum_{r=l}^{m}\binom{m}{r}q^r(1-q)^{m-r} = I_q(l, m-l+1). \tag{3.41}$$

Proof. We first derive (3.37). We start by using an expression for the expectation of a nonnegative RV with support $\{c^r, (c+1)^r, \ldots\}$ in terms of its complementary CDF:

$$\begin{aligned}
\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] &= c^r + \sum_{m=c^r}^{\infty}\mathbb{P}\left(\left(Y_{n:n}^{(c)}\right)^r > m\right) \\
&= c^r + \sum_{m=c^r}^{\infty}\mathbb{P}\left(Y_{n:n}^{(c)} > \lfloor m^{\frac{1}{r}}\rfloor\right) \\
&= c^r + \sum_{m=c^r}^{\infty}\left(1 - \mathbb{P}\left(Y_{n:n}^{(c)} \leq \lfloor m^{\frac{1}{r}}\rfloor\right)\right) \\
&= c^r + \sum_{m=c^r}^{\infty}\left(1 - \prod_{j=1}^{n}\mathbb{P}\left(Y_j^{(c)} \leq \lfloor m^{\frac{1}{r}}\rfloor\right)\right)
\end{aligned} \tag{3.42}$$

Next, fix $j \in [n]$ and observe $\{Y_j^{(c)} \leq \lfloor m^{\frac{1}{r}}\rfloor\}$ means that there are c successes in the first t trials, for some $t \in \{c, c+1, \ldots, \lfloor m^{\frac{1}{r}}\rfloor\}$. Conditioning on t, the c successes occurred over trials $\{1,\ldots,t\}$ with the number of trials between successive successes given by a c-vector $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_c)$ such that $\alpha_k \in \mathbb{N}$ and $\alpha_1 + \cdots + \alpha_c = t$. Fixing the particular sequence of successes and failures over trials $\{1,\ldots,t\}$, that sequence of outcomes has probability $\prod_{k\in[c]}(1-q_{jk})^{\alpha_k - 1}q_{jk}$. This yields (3.37).

To obtain (3.39) from (3.37) set $q_{jk} = q_j$ and observe

$$\sum_{\boldsymbol{\alpha}\in\mathcal{A}_t}\prod_{k=1}^{c}(1-q_j)^{\alpha_k-1}q_j = \sum_{\boldsymbol{\alpha}\in\mathcal{A}_t} q_j^c(1-q_j)^{t-c} = q_j^c(1-q_j)^{t-c}|\mathcal{A}_t|, \tag{3.43}$$

where $|\mathcal{A}_t|$ is the cardinality of $\mathcal{A}_t$, i.e., the number of c-vectors of positive integers that sum to t.

It can be shown that $|\mathcal{A}_t| = \binom{t-1}{c-1}$; this yields (3.39).

To derive (3.40) which holds for the state-independent case, observe $Y_j^{(c)}$ is a negative binomial RV, i.e., the number of trials required to obtain c receptions where each trial yields a reception with probability q_j. Let $W_j \sim \text{Bin}(\lfloor m^{\frac{1}{r}} \rfloor, q_j)$ be a binomial RV counting the number of receptions in $\lfloor m^{\frac{1}{r}} \rfloor$ trials, each trial having reception probability q_j. Observe the equivalence of the events $\{Y_j^{(c)} > \lfloor m^{\frac{1}{r}} \rfloor\} = \{W_j < c\}$. Taking probabilities of both sides, and applying (3.41) yields:

$$\begin{aligned}
\mathbb{P}\left(Y_j^{(c)} > \lfloor m^{\frac{1}{r}} \rfloor\right) &= \mathbb{P}(W_j < c) = 1 - \mathbb{P}(W_j \geq c) \\
&= 1 - \sum_{t=c}^{\lfloor m^{\frac{1}{r}} \rfloor} \mathbb{P}(W_j = t) = 1 - \sum_{t=c}^{\lfloor m^{\frac{1}{r}} \rfloor} \binom{\lfloor m^{\frac{1}{r}} \rfloor}{t} q_j^t (1-q_j)^{\lfloor m^{\frac{1}{r}} \rfloor - t} \\
&= 1 - I_{q_j}(c, \lfloor m^{\frac{1}{r}} \rfloor - c + 1) \qquad (3.44)
\end{aligned}$$

Starting from (3.42) and substituting (3.44) yields (3.40). □

3.4.3 Recurrence for the (moments of) delay under RLC

A limitation of Prop. 9 is that the expression involves an infinite summation. In this subsection, we offer a recurrence equation for the RV $Y_{n:n}$ that permits calculation of the exact value for $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ using only a finite number of terms. In fact it is convenient to generalize our setting slightly, in that we now suppose that receiver j requires reception of c_j packets; previous to this we have assumed $c_j = c$ for all $j \in [n]$. Let $\mathbf{c}^0 = (c_1^0, \ldots, c_n^0)$ be the n-vector of required number of successes for each receiver. The recurrence will be in terms of the generic vector $\mathbf{c} \leq \mathbf{c}^0$ (component-wise), interpreted as the number of successes left to go for each receiver, as explained below. We shall also in this subsection write $Y_{n:n}^{(\mathbf{c})}$ to emphasize this.

We introduce some shorthand notation. First, define $[n(\mathbf{c})] \equiv \{j \in [n] : c_j > 0\}$ as the set of active receivers, i.e., those still requiring an additional reception to complete. Second, define the probabilities of success and failure by the active receiver subsets S and $[n(\mathbf{c})] \setminus S$ respectively, to be:

$$q(S, \mathbf{c}) \equiv \prod_{j \in S} q_{j, c_j^0 - c_j + 1}, \quad \bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}) \equiv \prod_{j \in [n(\mathbf{c})] \setminus S} (1 - q_{j, c_j^0 - c_j + 1}). \qquad (3.45)$$

In words, $q(S, \mathbf{c})$ is the probability of success for active receivers with indices in S where the "successes to go" c_j and initial number of successes required c_j^0 determine the state index $k = c_j^0 - c_j + 1$ for receiver j. Further, $\bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c})$ is the probability of failure for the active receivers not indexed by S, where again the state index for each such receiver j is $k = c_j^0 - c_j + 1$.

Third, define $\mathbf{1}_S$ to be the n-vector with ones in the positions indexed by S and zero elsewhere. We reiterate that in the state-dependent case with success probability matrix $\mathbf{Q}$, we have $Y_{n:n}^{(\mathbf{c})} = \max(Y_1^{(c_1)}, \ldots, Y_n^{(c_n)})$ where $(Y_1^{(c_1)}, \ldots, Y_n^{(c_n)})$ are independent with $Y_j^{(c_j)} = \sum_{k=1}^{c_j} X_{j,k}$ and the $X_{j,k} \sim \text{Geo}(q_{j,k})$ are themselves independent in j and k. In the state-independent case we have $X_{j,k} \sim \text{Geo}(q_j)$ for $k \in [c_j]$.

Proposition 10. *Assume A1: State-dependent receptions, heterogeneous receivers. The RV* $Y_{n:n}^{(\mathbf{c}^0)}$ *defined above admits the recurrence*

$$Y_{n:n}^{(\mathbf{c})} = 1 + \sum_{S \subseteq [n(\mathbf{c})]} q(S, \mathbf{c}) \bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}) Y_{n:n}^{(\mathbf{c} - \mathbf{1}_S)}, \tag{3.46}$$

with boundary condition $Y_{n:n}^{(\mathbf{0})} = 0$. *Assume A2: State-independent receptions, heterogeneous receivers. The RV* $Y_{n:n}^{(\mathbf{c}^0)}$ *defined above admits the recurrence*

$$Y_{n:n}^{(\mathbf{c})} = 1 + \sum_{S \subseteq [n(\mathbf{c})]} q(S, \mathbf{c}^0) \bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}^0) Y_{n:n}^{(\mathbf{c} - \mathbf{1}_S)}, \tag{3.47}$$

again with boundary condition $Y_{n:n}^{(\mathbf{0})} = 0$.

Proof. The recurrence is obtained by conditioning on the outcomes of the next trial starting from the given "state" $\mathbf{c}^0$ indicating the number of successes required by each receiver. The set of outcomes for the first trial is all subsets S of the active receivers $[n(\mathbf{c})]$, i.e., all possible subsets of potential successes. The probability of success by active receivers in S and failure by active receivers not in S is $q(S, \mathbf{c}) \bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c})$. The effect of successes by receivers in S is to reduce the number of required succeses by those receivers by one, i.e., $\mathbf{c} \to \mathbf{c} - \mathbf{1}_S$, thus advancing the active receivers in S to the next "state". That is, their success probability advances from q_{j,k_j} where $k_j = c_j^0 - c_j + 1$ to q_{j,k_j+1}.

This gives (3.46). To obtain (3.47), observe that in the state-independent case, the probability of success by S and failure by $[n(\mathbf{c})] \setminus S$ is $\prod_{j \in S} q_j$ times $\prod_{j \in [n(\mathbf{c})] \setminus S} (1 - q_j)$, which is the same as $q(S, \mathbf{c}^0)\bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}^0)$. □

Remark 11. *The recurrence hinges critically on the fact that the $X_{j,k}$ are geometric RVs, i.e., on the memoryless property. Conditioning on the number of successes affects the state $\mathbf{c}$ but, aside from that, the system probabilistically restarts itself.*

Remark 12. *The empty set is included in the set of subsets of $[n(\mathbf{c})]$ in* (3.46) *and* (3.47)*, which must be "subtracted out" to express the recurrence for $Y_{n:n}^{(\mathbf{c})}$ in terms of strictly smaller vectors $\mathbf{c}' < \mathbf{c}$ (in at least one component). Doing this gives, for* (3.46)*,*

$$Y_{n:n}^{(\mathbf{c})} = \frac{1 + \sum_{S \subseteq [n(\mathbf{c})] \setminus \emptyset} q(S, \mathbf{c})\bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}) Y_{n:n}^{(\mathbf{c} - \mathbf{1}_S)}}{1 - q(\emptyset, \mathbf{c})\bar{q}([n(\mathbf{c})], \mathbf{c})}, \tag{3.48}$$

where $[n(\mathbf{c})] \setminus \emptyset$ denotes all non-empty subsets of the set of active receivers.

Taking the expectations of (3.46) and (3.47) yields recurrences on $\mathbb{E}[Y_{n:n}^{(c)}]$ for the state dependent and independent cases.

Corollary 3. *Assume A1: State-dependent receptions, heterogeneous receivers. The expectation $\mathbb{E}[Y_{n:n}^{(\mathbf{c})}]$ admits the recurrence*

$$\mathbb{E}\left[Y_{n:n}^{(\mathbf{c})}\right] = 1 + \sum_{S \subseteq [n(\mathbf{c})]} q(S, \mathbf{c})\bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}) \mathbb{E}\left[Y_{n:n}^{(\mathbf{c} - \mathbf{1}_S)}\right], \tag{3.49}$$

with boundary condition $\mathbb{E}[Y_{n:n}^{(\mathbf{0})}] = 0$. Assume A2: State-independent receptions, heterogeneous receivers. The expectation $\mathbb{E}[Y_{n:n}^{(\mathbf{c})}]$ admits the recurrence

$$\mathbb{E}\left[Y_{n:n}^{(\mathbf{c})}\right] = 1 + \sum_{S \subseteq [n(\mathbf{c})]} q(S, \mathbf{c}^0)\bar{q}([n(\mathbf{c})] \setminus S, \mathbf{c}^0) \mathbb{E}\left[Y_{n:n}^{(\mathbf{c} - \mathbf{1}_S)}\right], \tag{3.50}$$

again with boundary condition $\mathbb{E}[Y_{n:n}^{(\mathbf{0})}] = 0$.

In fact it is straightforward to see that for $\mathbf{c} = c_j \mathbf{e}_j$ (the n-vector of all zeros with value c_j in

position j), in the state-independent case we have the boundary condition $\mathbb{E}[Y_{n:n}^{(c_j\mathbf{e}_j)}] = c_j/q_j$. That is, when there is only one receiver left to receive, the expected duration is the expectation of a negative binomial RV, which of course is c_j succcesses required times $1/q_j$ time slots on average per success.

Remark 13. *A parallel recurrence holds for the RV* $Y_{1:n} = Y_{1:n}^{(\mathbf{c}^0)}$ *where* $Y_{1:n}^{(\mathbf{c}^0)} = \min(Y_1^{(c_1^0)}, \ldots, Y_n^{(c_n^0)})$ *is the random time until the* first *receiver, say* j*, receives its target number of successes* c_j^0 *(as opposed to counting until all receivers reach their targets). For the state-dependent case the RV* $Y_{1:n} = Y_{1:n}^{(\mathbf{c}^0)}$ *defined above admits the recurrence*

$$Y_{1:n}^{(\mathbf{c})} = 1 + \sum_{S\subseteq[n]} q(S,\mathbf{c})\bar{q}([n]\setminus S,\mathbf{c})Y_{1:n}^{(\mathbf{c}-\mathbf{1}_S)}, \tag{3.51}$$

with boundary condition $Y_{1:n}^{(\mathbf{c})} = 0$ *for any* $\mathbf{c}$ *containing one or more zeros. For the state-independent case the RV* $Y_{1:n} = Y_{1:n}^{(\mathbf{c}^0)}$ *defined above admits the recurrence*

$$Y_{1:n}^{(\mathbf{c})} = 1 + \sum_{S\subseteq[n]} q(S,\mathbf{c}^0)\bar{q}([n]\setminus S,\mathbf{c}^0)Y_{1:n}^{(\mathbf{c}-\mathbf{1}_S)}, \tag{3.52}$$

again with boundary condition $Y_{1:n}^{(\mathbf{c})} = 0$ *for any* $\mathbf{c}$ *containing one or more zeros.*

Remark 14. *In fact we can use the memoryless property to establish recurrences to compute arbitrary moments of* $Y_{n:n}$*, i.e.,* $\mathbb{E}\left[\left(Y_{n:n}^{(\mathbf{c}^0)}\right)^r\right]$*. To see this, by the memoryless property:*

$$\mathbb{P}(Y_{n:n}^{(\mathbf{c})} = y \mid S \in [n(\mathbf{c})] \text{ succeed in c.t.s.}) = \mathbb{P}(1 + Y_{n:n}^{(\mathbf{c}-\mathbf{1}_S)} = y). \tag{3.53}$$

where "c.t.s." stands for current time slot*. That is, the event* $\{Y_{n:n}^{(\mathbf{c})} = y\}$ *conditioned on the event of successes for active receivers* S *in the current time slot, is the same as the event* $\{1 + Y_{n:n}^{(\mathbf{c}-\mathbf{1}_S)} = y\}$*, which is just an example of the common first-step analysis technique in probability. But this allows us to equate arbitrary powers* r *of both sides, i.e.,*

$$\mathbb{P}\left(\left(Y_{n:n}^{(\mathbf{c})}\right)^r = y^r \mid S \in [n(\mathbf{c})] \text{ succeed in c.t.s}\right) = \mathbb{P}\left(\left(1 + Y_{n:n}^{(\mathbf{c}-\mathbf{1}_S)}\right)^r = y^r\right). \tag{3.54}$$

On this basis, our recurrence can be generalized to one for higher moments of $Y_{n:n}$, e.g., for the state-dependent case:

$$\mathbb{E}\left[\left(Y_{n:n}^{(\mathbf{c})}\right)^r\right] = \sum_{S\subseteq[n(\mathbf{c})]} q(S,\mathbf{c})\bar{q}([n(\mathbf{c})]\setminus S,\mathbf{c})\mathbb{E}\left[\left(1+Y_{n:n}^{(\mathbf{c}-\mathbf{1}_S)}\right)^r\right], \tag{3.55}$$

with boundary condition $\mathbb{E}\left[\left(Y_{n:n}^{(\mathbf{0})}\right)^r\right] = 0$.

Remark 15. *Recurrences for maximum of geometric RVs and in fact maximum of sums of geometric RVs are found in the literature. We mention in particular[44] (1990 (2.5)) which provided the inspiration for our recurrence, but in the simpler context of the expected maximum of iid geometric RVs. Further, recent work[45] (2010) on the expected delay of network coded packets routed simultaneously over two paths employed a similar recurrence to ours, but with significant differences. In particular, since in their context there is a single receiver looking for c innovative packets over $n = 2$ disjoint routes, their recurrence is univariate in the scalar c. The general problem of expressing the maximum of negative binomial RVs is addressed in[46].*

3.4.4 Erasure channels that minimize delay

In this subsection we are interested in characterizing the channels for which $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$ is minimized, subject to a constraint on the average over the n channel reception probabilities. Throughout this subsection we restrict our attention to the case $q_{j,k} = q_j$ for all $j \in [n]$, or equivalently, we assume the field size d is infinite in (3.1). Denote the corresponding channel reception probabilities as $\mathbf{q} = (q_j, j \in [n])$, the average as $\bar{\mathbf{q}} = (1/n)\sum_{j=1}^n q_j$, and the set of possible channel vectors with average q as $\mathcal{Q}_q = \{\mathbf{q} \in (0,1)^n : \bar{\mathbf{q}} = q\}$, for $q \in (0,1)$. Prop. 11 (12) establish that the delay minimizing channel vector is $\mathbf{q} = (q,\ldots,q)$ when minimized over $\mathcal{Q}_q$, for UT (RLC), respectively. As RLC includes UT as a special case under Assumption A2 (state-independent receptions, heterogeneous receivers), Prop. 12 implies Prop. 11. We include both because the proof of 11 is substantially simpler than that of 12 while still retaining the key ideas.

Proposition 11. *For any $q \in (0,1)$, the r^{th} ($r \in \mathbb{N}$) moment of delay under UT, $\mathbb{E}[X_{n:n}^r]$, is minimized over $\mathbf{q} \in \mathcal{Q}_q$ for $\mathbf{q} = (q,\ldots,q)$.*

Proof. Write $q\mathbf{1}$ to denote $(q,\ldots,q)$. We need to show $\mathbb{E}[X^r_{n:n}(\mathbf{q})] \geq \mathbb{E}[X^r_{n:n}(q\mathbf{1})]$ for all $\mathbf{q} \in \mathcal{Q}_q$. Towards this, we write

$$\begin{aligned}
\mathbb{E}[X^r_{n:n}] &= \sum_{x=0}^{\infty} \mathbb{P}(X^r_{n:n} > x) \\
&= \sum_{x=0}^{\infty} \mathbb{P}(X_{n:n} > \lfloor x^{\frac{1}{r}} \rfloor) \\
&= \sum_{x=0}^{\infty} \left(1 - \prod_{j\in[n]} \mathbb{P}(X_j \leq \lfloor x^{\frac{1}{r}} \rfloor)\right) \\
&= 1 + \sum_{x=1}^{\infty} \left(1 - \prod_{j\in[n]} \left(1 - (1-q_j)^{\lfloor x^{\frac{1}{r}} \rfloor}\right)\right). \qquad (3.56)
\end{aligned}$$

It suffices to show the desired inequality holds termwise with respect to the summing variable x:

$$1 - \prod_{j\in[n]} \left(1 - (1-q_j)^{\lfloor x^{\frac{1}{r}} \rfloor}\right) \geq 1 - \left(1 - (1-q)^{\lfloor x^{\frac{1}{r}} \rfloor}\right)^n, \; x \in \mathbb{N}. \qquad (3.57)$$

This is equivalent to showing

$$\frac{1}{n} \sum_{j\in[n]} \log\left(1 - (1-q_j)^{\lfloor x^{\frac{1}{r}} \rfloor}\right) \leq \log\left(1 - (1-q)^{\lfloor x^{\frac{1}{r}} \rfloor}\right). \qquad (3.58)$$

Given x, define $f(z) \equiv \log(1-(1-z)^{\lfloor x^{\frac{1}{r}} \rfloor})$. Further, define a discrete uniform random variable Z with support $\{q_1,\ldots,q_n\}$. By assumption $\mathbb{E}[Z] = \bar{\mathbf{q}} = q$. Consequently (3.58) directly follows from Jensen's inequality and the concavity of $f(z)$ on $z \in (0,1)$, for each $x \in \mathbb{N}$. □

Proposition 12. *Assume A2: State-independent receptions, heterogeneous receivers. For any $c \in \mathbb{N}$ and $q \in (0,1)$, the r^{th} ($r \in \mathbb{N}$) moment of delay under RLC, $\mathbb{E}\left[\left(Y^{(c)}_{n:n}\right)^r\right]$, is minimized over $\mathbf{q} \in \mathcal{Q}_q$ for $\mathbf{q} = (q,\ldots,q)$.*

Proof. The proof mirrors that of Prop. 11. Write $q\mathbf{1}$ to denote $(q,\ldots,q)$. We need to show $\mathbb{E}\left[\left(Y^{(c)}_{n:n}(\mathbf{q})\right)^r\right] \geq \mathbb{E}\left[\left(Y^{(c)}_{n:n}(q\mathbf{1})\right)^r\right]$ for all $\mathbf{q} \in \mathcal{Q}_q$. An expression for $\mathbb{E}\left[\left(Y^{(c)}_{n:n}\right)^r\right]$ in terms of regularized incomplete Beta function is given in §3.4.2 Eq. (3.40). To establish this inequality, it suffices to show this inequality holds for each term in the sum. That is, for each integer $m \geq c^r$, we

must show

$$1-\prod_{j=1}^{n} I_{q_j}\left(c, \lfloor m^{\frac{1}{r}} \rfloor - c + 1\right) \geq 1 - \left(I_q\left(c, \lfloor m^{\frac{1}{r}} \rfloor - c + 1\right)\right)^n. \tag{3.59}$$

Since the function $I_x(a,b)$ is always nonnegative, the above is equivalent to

$$\frac{1}{n}\sum_{j=1}^{n} \log I_{q_j}(c, \lfloor m^{\frac{1}{r}} \rfloor - c + 1) \leq \log I_q(c, \lfloor m^{\frac{1}{r}} \rfloor - c + 1). \tag{3.60}$$

Given integer $m \geq c^r$, define $f(z) \equiv \log I_z(c, \lfloor m^{\frac{1}{r}} \rfloor - c + 1)$, which is shown in the Appendix (§3.7.1, Lem. 4) to be concave on $z \in (0,1)$. Then (3.60) follows from Jensen's inequality where the associated RV is again a discrete uniform with support $\{q_1, \ldots, q_n\}$. □

Remark 16. *A similar argument establishes the fact that replacing the success probabilities of any subset of receivers with the average over that subset will yield a lower average delay. Specifically, for any* $\mathbf{q} \in (0,1)^n$ *and any* $\mathcal{S} \subseteq [n]$*, form* $\mathbf{q}'$ *with* $q'_j = q_j$ *for* $j \notin \mathcal{S}$ *and* $q'_j = \sum_{i\in\mathcal{S}} q_i/|\mathcal{S}|$ *for* $j \in \mathcal{S}$*. Then* $\mathbb{E}\left[\left(Y^{(c)}_{n:n}\right)^r\right]$ *is no larger under* $\mathbf{q}'$ *than under* $\mathbf{q}$*.*

Remark 17. *As will be shown in Prop. 13, the asymptotic delay per packet (as* $c \to \infty$*) converges to* $1/\min_{j\in[n]} q_j$*, meaning the asymptotic delay depends upon the channel vector* $\mathbf{q}$ *only through its minimum value (bottleneck channel). Given this, Prop. 12 is not surprising since the maximizer of* $\min \mathbf{q}$ *over* $\mathbf{q} \in \mathcal{Q}_q$ *is* $\mathbf{q} = (q, \ldots, q)$*:*

$$\arg\min_{\mathbf{q}\in\mathcal{Q}_q} \frac{1}{\min_{j\in[n]} q_j} = \arg\max_{\mathbf{q}\in\mathcal{Q}_q} \min_{j\in[n]} q_j = (q, \ldots, q). \tag{3.61}$$

We emphasize, however, that Prop. 12 holds for all $c \in \mathbb{N}$*, whereas Prop. 13 holds as* $c \to \infty$*.*

3.5 Dependence of RLC delay on the blocklength c

In the previous section we derived exact expressions (involving infinite sums) as well as a recurrence for the moments of the RLC block delay. Although these expressions have the virtue of being exact and computable, they fail to provide insight on the delay's dependence on the key model parameters c, n. This and the subsequent section address this deficiency by studying the asymptotic

(occasionally normalized) block delay as c or n grows to infinity. In this section we investigate the dependence on c with n fixed, while in the next section (§3.6) we investigate the dependence on n with c fixed. This section contains several subsections. First, we show in §3.5.1 the delay per packet converges in probability and in r^{th} mean. This result allows us to easily establish that the asymptotic (in c) delay per packet under RLC is lower than the packet delay under UT. Second, we show in §3.5.2 a stronger result (implying the previous result) that the expected delay per packet under RLC is monotone decreasing in c for both infinite and finite field sizes. Furthermore, the monotonicity properties are investigated from a sample path perspective. Third, we show in §3.5.3 that the lower and upper bounds in §3.4.1 are (almost) asymptotically tight as $c \to \infty$. Throughout this section we employ the superscript (c) on all quantities (not just $Y_{n:n}^{(c)}, Y_j^{(c)}$) to emphasize the dependence upon c.

3.5.1 Asymptotic delay per packet as $c \to \infty$

Recall the $n \times c$ matrix $\mathbf{Q}^{(c)}$ holds the parameters $q_{j,k}^{(c)}$, where $X_{j,k}^{(c)} \sim \text{Geo}(q_{j,k}^{(c)})$. The first main result (Prop. 13) establishes that, under some convergence requirements on $\{X_{j,k}^{(c)}\}$, the asymptotic per packet delay $(Y_{n:n}^{(c)}/c)$ converges to $1/\min_j q_j$ as $c \to \infty$ in both probability and r^{th} mean, where q_j is the asymptotic (in c) average of row j of $\mathbf{Q}^{(c)}$. That is, the asymptotic packet delay depends solely on the reception probability of the bottleneck channel(s), $\min_{j\in[n]} q_j$, and is in fact independent of n. The conditions required in Prop. 13 are in fact satisfied for the specific $q_{j,k}^{(c)}$ in (3.1) capturing the effect of the field size on the probability of linear independence (Prop. 14). The second main result (Prop. 15) uses this to conclude the asymptotic per packet delay of RLC is lower than the packet delay under UT. Normalize $(Y_1^{(c)}, \ldots, Y_n^{(c)})$ as $(\hat{Y}_1^{(c)}, \ldots, \hat{Y}_n^{(c)})$ with

$$\hat{Y}_j^{(c)} \equiv \frac{Y_j^{(c)}}{c} = \frac{1}{c}\sum_{k=1}^{c} X_{j,k}^{(c)},\ j \in [n].$$

Similarly, normalize $Y_{n:n}^{(c)}$ as $\hat{Y}_{n:n}^{(c)} \equiv Y_{n:n}^{(c)}/c$.

Proposition 13. *Assume A1: State-dependent receptions, heterogeneous receivers. Given the $n \times c$ matrix $\mathbf{Q}^{(c)}$ with success probabilities $q_{j,k}^{(c)}$, suppose there exist n-vectors $\mathbf{q} = (q_1, \ldots, q_n)$ and $\boldsymbol{\sigma}^2 =$*

$(\sigma_1^2, \ldots, \sigma_n^2)$ such that the following hold for all $j \in [n]$:

$$\begin{aligned} \lim_{c\to\infty} \frac{1}{c} \sum_{k=1}^{c} \mathbb{E}[X_{j,k}^{(c)}] &= \frac{1}{q_j} < \infty \\ \lim_{c\to\infty} \frac{1}{c} \sum_{k=1}^{c} \mathrm{Var}(X_{j,k}^{(c)}) &= \sigma_j^2 < \infty \end{aligned} \tag{3.62}$$

Then, as $c \to \infty$, $\hat{Y}_{n:n}^{(c)}$ converges in probability:

$$\hat{Y}_{n:n}^{(c)} \xrightarrow{\mathbb{P}} (\min_{j\in[n]} q_j)^{-1}. \tag{3.63}$$

Further, fix $r \in \mathbb{N}$ and suppose that each $X_{j,k}^{(c)}$ has uniformly bounded moment up to order $2r$. Then (3.62) *along with this assumption ensure that, as $c \to \infty$, $\hat{Y}_{n:n}^{(c)}$ converges in r^{th} mean:*

$$\hat{Y}_{n:n}^{(c)} \xrightarrow{L^r} (\min_{j\in[n]} q_j)^{-1}. \tag{3.64}$$

Proof. We first provide the outline of the proof. To establish convergence in probability of $\hat{Y}_{n:n}^{(c)}$ we first establish convergence in probability of each component of the vector $(\hat{Y}_1^{(c)}, \ldots, \hat{Y}_n^{(c)})$ via the weak law of large numbers (WLLN) for independent but not necessarily identically distributed RVs. Convergence in probability of each component of a sequence of vector-valued random variables to a limit ensures convergence in probability of the vector as a whole. Next, use the fact that convergence in probability is preserved under continuous functions, and in particular the function $\max(y_1, \ldots, y_n)$. This establishes convergence in probability of $\hat{Y}_{n:n}^{(c)}$.

Furthermore, to establish convergence in r^{th} mean of $\hat{Y}_{n:n}^{(c)}$, we first use the fact that if there exists a finite constant, say R, such that the second moment of each element of the sequence is uniformly bounded by R, then the sequence itself is *uniformly integrable* (UI). Then, we employ the fact that a sequence of RVs that converges in probability and is UI with parameter r must converge in r^{th} mean.

We now establish convergence in probability. Prop. 14 shows (3.62) is satisfied, then according to the WLLN for independent but not necessarily identically distributed RVs (e.g.,[47] Theorem 1.1

in §7), we have:

$$\hat{Y}_j^{(c)} \xrightarrow{\mathbb{P}} \mathbb{E}\left[\lim_{c\to\infty} \frac{1}{c}\sum_{k=1}^{c} X_{jk}^{(c)}\right] = \frac{1}{q_j},\ j \in [n]. \tag{3.65}$$

Convergence in probability of each component $j \in [n]$ ensures convergence in probability of the vector as a whole:

$$\left(\hat{Y}_1^{(c)}, \ldots, \hat{Y}_n^{(c)}\right) \xrightarrow{\mathbb{P}} \left(\frac{1}{q_1}, \ldots, \frac{1}{q_n}\right). \tag{3.66}$$

Since convergence in probability is preserved under continuous functions such as $\max(y_1, \ldots, y_n)$, it follows that:

$$\hat{Y}_{n:n}^{(c)} = \max\left(\hat{Y}_1^{(c)}, \ldots, \hat{Y}_n^{(c)}\right) \xrightarrow{\mathbb{P}} \max\left(\frac{1}{q_1}, \ldots, \frac{1}{q_n}\right). \tag{3.67}$$

This establishes convergence in probability of $\hat{Y}_{n:n}^{(c)}$ in c to $1/\min_j q_j$.

We next establish convergence in r^{th} mean. We first establish the sequence $\left(\hat{Y}_{n:n}^{(c)}\right)$ in c is uniformly integrable for each $r \in \mathbb{N}$. A sufficient condition for this is to show there exists a constant R such that $\mathbb{E}\left[\left(\left(\hat{Y}_{n:n}^{(c)}\right)^r\right)^2\right] < R$ for all c. We establish this as follows:

$$\begin{aligned}
\left(\hat{Y}_{n:n}^{(c)}\right)^r &= \left(\max\left(\hat{Y}_1^{(c)}, \ldots, \hat{Y}_n^{(c)}\right)\right)^r \\
&= \max\left(\left(\hat{Y}_1^{(c)}\right)^r, \ldots, \left(\hat{Y}_n^{(c)}\right)^r\right) \\
&\leq \sum_{j=1}^{n}\left(\hat{Y}_j^{(c)}\right)^r = \sum_{j=1}^{n}\left(\frac{1}{c}\sum_{k=1}^{c} X_{j,k}^{(c)}\right)^r.
\end{aligned} \tag{3.68}$$

We now establish the sequence of RVs $\left(\hat{Y}_{n:n}^{(c)}\right)^r$ is UI. It suffices ([47] Thm. 4.2 and Remark 4.3 in §7) to show $\mathbb{E}\left[\left(\hat{Y}_{n:n}^{(c)}\right)^{2r}\right] < R$ for all c.

$$\mathbb{E}\left[\left(\hat{Y}_{n:n}^{(c)}\right)^{2r}\right] \leq \mathbb{E}\left[\sum_{j=1}^{n}\left(\frac{1}{c}\sum_{k=1}^{c} X_{j,k}^{(c)}\right)^{2r}\right]. \tag{3.69}$$

It follows that there exists a constant $R < \infty$ such that the RHS of (3.69) is bounded above by R provided each $X_{j,k}^{(c)}$ has uniformly bounded moments up to order $2r$. Next, it is a theorem ([47] Thm. (4.1) in §7) that convergence in probability and uniform integrability with parameter r ensures convergence in r^{th} mean. □

Proposition 14. *Assume A1: State-dependent receptions, heterogeneous receivers. Conditions* (3.62) *of Prop. 13 on the mean and variance of* $\{X_{j,k}^{(c)}\}$ *for the weak law of large numbers to apply are satisfied for* $q_{j,k}^{(c)} = (1 - d^{k-1-c})q_j$ *as in* (3.1)*, for any field size* $d \geq 2$*.*

The proof may be found in the Appendix.

Remark 18. *In particular for* $r = 1$ *it easily follows from Prop. 13 that* $\mathbb{E}[\hat{Y}_{n:n}^{(c)}] \to 1/\min_j q_j$ *as* $c \to \infty$*. Note*[33] *establishes this fact as well, but by quite different means. Namely, they derive upper and lower bounds on* $\mathbb{E}[\hat{Y}_{n:n}^{(c)}]$ *for each* c *and show that these bounds converge in* c *to* $1/\min_j q_j$*.*

Remark 19. *Prop. 13 reveals something important about RLC over the erasure broadcast channel. Namely, since* $\mathbb{E}[\hat{Y}_{n:n}^{(c)}] \to \left(\min_{j\in[n]} q_j\right)^{-1}$ *depends upon* $\mathbf{q}$ *(and thus* n*) only through the statistic* $\min_{j\in[n]} q_j$*. Thus, in particular, the average delay per packet for* n *channels with success probabilities* $\mathbf{Q}^{(c)}$ *is asymptotically in* c *the same as the average delay per packet of a single erasure channel (i.e.,* $n' = 1$*) with* $q_1 = \min_{j\in[n]} q_j$*. In short, the performance of the broadcast erasure channel is asymptotically (in* c*) limited by the bottleneck channel.*

Proposition 15. *Assume A1: State-dependent receptions, heterogeneous receivers. The asymptotic expected delay per packet under RLC is superior to the expected delay per packet under UT:*

$$\lim_{c\to\infty} \frac{\mathbb{E}[Y_{n:n}^{(c)}]}{c} = \frac{1}{\min_{j\in[n]} q_j} \leq \mathbb{E}[X_{n:n}]. \tag{3.70}$$

Proof. Fix the number of receivers n and the reception probabilities $\mathbf{q} = (q_1, \ldots, q_n)$. Suppose $i \in \arg\min_{j\in[n]} q_j$ and fix an arbitrary other index $k \in [n] \setminus i$. Consider two copies of this channel, one running RLC with $c \to \infty$ for all n receivers with reception probabilities $\mathbf{q}$, and the other running UT on with only the pair of 2 receivers, i, k, selected from $[n]$, with reception probabilities $0 < q_i \leq q_k$. The expected delay per packet under RLC is $1/q_i$, and the expected delay per packet under UT on the pruned systems with 2 receivers is, using Cor. 2,

$$\mathbb{E}[\max(X_i, X_k)] \quad = \quad (-1)^{1+1}\left(\frac{1}{1-(1-q_i)} + \frac{1}{1-(1-q_k)}\right) + (-1)^{2+1}\frac{1}{1-(1-q_i)(1-q_k)}.$$

Observe that if we can show $1/\min_j q_j \leq \mathbb{E}[X_{n:n}]$ for the pruned system, then clearly it also holds for the original system with all n receivers. Simple algebra establishes that $1/q_i \leq \mathbb{E}[\max(X_i, X_k)]$, where the inequality is tight only in the degenerate cases when either $q_i = 0$ or $q_k = 1$. □

Remark 20. *The above proof attempted using the lower bound $\psi_1(\mathbf{q})$ from Prop. 3, instead of the exact expression for $\mathbb{E}[X_{n:n}]$ from Prop. 4 would fail as there are non-degenerate choices for $\mathbf{q}$ for which the lower bound on expected delay per packet under UT could be superior to that of RLC. This is in fact a key motivation for developing exact expressions for the moment of $X_{n:n}$ (Prop. 4). The desired inequality does go through when using the upper bound on $\mathbb{E}[X_{n:n}]$, namely $\psi_1(\mathbf{q}) + 1$, from Prop. 3, but this inequality is inconclusive as it only relates an upper bound on UT to the asymptotic performance of RLC.*

3.5.2 Monotonicity properties of delay per packet

The previous subsection demonstrates the advantage in expected delay per packet of RLC over UT in the asymptotic regime, as the blocklength $c \to \infty$. In this subsection we establish certain monotonicity properties for the delay per packet under RLC as a function of c, with the number of receivers n held fixed. We provide four results. Prop. 16 and 17 establish the expected delay per packet is monotone decreasing in the blocklength when the field size is infinite and finite, respectively. Thus, given two blocklengths c_1 and c_2, respectively, the above propositions give that the expected delay per packet for under c_1 is less than that under c_2 if $c_1 > c_2$. In contrast, Props. 18 and 19 establish "negative" results. Prop. 18 establishes the sequence of random delay per packet at each receiver j, $\{Y_j^{(c)}/c\}_c$, is not stochastically ordered, while Prop. 19 establishes that in order for the delay per packet under c_1 to be no larger than that under c_2 for all sample path realizations, it is necessary and sufficient that $c_1 = kc_2$ for some integer $k \geq 2$. The proofs of Prop. 16, 17, and 19 are in the Appendix.

Consider a collection of RVs $(X_{j,k}^{(c)},\ j \in [n], k \in [c])$ where $X_{j,k}^{(c)} \sim \text{Geo}(q_{j,k}^{(c)})$ represents the time required, after obtaining the $(k-1)^{\text{st}}$ linearly independent packet encoding, to obtain the k^{th} linearly independent packet encoding by receiver j. Recall our assumption that $X_{j,k}$ are independent in j

and k. Further recall $Y_j^{(c)} = X_{j,1}^{(c)} + \cdots + X_{j,c}^{(c)}$ for $j \in [n]$ represents the decoding delay for each receiver j to recover all c packets. Define the normalized expected maximum delay difference as

$$\Delta^{(c)} \equiv \frac{1}{c}\mathbb{E}\left[Y_{n:n}^{(c)}\right] - \frac{1}{c+1}\mathbb{E}\left[Y_{n:n}^{(c+1)}\right]. \tag{3.71}$$

Note that $\{\Delta^{(c)}\}$ is positive for all c if and only if $\{\mathbb{E}[Y_{n:n}^{(c)}]/c\}$ is monotone decreasing in c, and we prove monotonicity by establishing the sign of $\Delta^{(c)}$ for each c.

In the proofs that follow it will be of use to characterize the random set of indices $\mathcal{J}^{(c)} = \arg\max(Y_j^{(c)}, j \in [n])$ with the largest decoding delay. Equivalently, $\mathcal{J}^{(c)}$ is the set of row indices in the random matrix $\mathbf{X}^{(c)} \equiv (X_{j,k}^{(c)}, j \in [n], k \in [c])$ with the largest row-sum. More generally, we are interested in matrix row selection rules that are column-invariant, as is the max row-sum selection rule. Define a row-index selection rule $\chi : \mathbb{R}^{n\times c} \to [n]$ that produces a row index $j = \chi(A)$ for any matrix A, and in particular returns a random variable $J = \chi(\mathbf{X}^{(c)})$ for the random matrix $\mathbf{X}^{(c)}$. Let σ be any permutation of $[c]$ and $\sigma(\mathbf{X}^{(c)})$ is the matrix $\mathbf{X}^{(c)}$ with its columns permuted according to σ. Say χ is column invariant if $\chi(\mathbf{X}^{(c)}) = \chi(\sigma(\mathbf{X}^{(c)}))$ for all $\sigma \in \Sigma$, where Σ is the set of all permutations of $[c]$. In particular we will be using the column invariant selection rule $\hat{J}^{(c)} = \min \hat{\mathcal{J}}^{(c)}$. In words, $\hat{J}^{(c)}$ is the smallest index in the (random) set of indices in $[n]$ that achieve the maximum $Y_j^{(c)}$. It would work as well to select any other element from $\hat{\mathcal{J}}^{(c)}$; the important point is that the selection $\hat{J}^{(c)}$ is uniquely determined for each $\hat{\mathcal{J}}^{(c)}$.

Proposition 16. *Assume A2: State-independent receptions, heterogeneous receivers. Then the expected delay per packet $\mathbb{E}[Y_{n:n}^{(c)}/c]$ is decreasing in the code blocklength c.*

Proposition 17. *Assume A1: State-dependent receptions, heterogeneous receivers. Then the expected delay per packet $\mathbb{E}[Y_{n:n}^{(c)}/c]$ is decreasing in the code blocklength c.*

Remark 21. *The fact that the row selection rule χ is column-invariant is leveraged in both of the proofs of the above two propositions. The key difference between them is that when the field size is infinite the matrix $\mathbf{X}^{(c)}$ has iid columns, whereas when the field size is finite the matrix $\mathbf{X}^{(c)}$ has independent columns with a column-dependent distribution. In fact, in the former case, the proof*

holds regardless of the distribution (hence our use of the generic RV Z in the proof), whereas in the latter case the proof requires properties of the assumed distribution for $\mathbf{X}^{(c)}$.

The following corollary illustrates a consequence of the above monotonicity propositions. Given a fixed number of M packets, any block coding based strategy must select a *block partition* $(m, \mathbf{c})$ consisting of *i*) the number of blocks $m \in \mathbb{N}$, and *ii*) block sizes $\mathbf{c} = (c_1, \ldots, c_m) \in \mathbb{N}^m$ satisfying $c_1 + \cdots + c_m = M$.

Corollary 4. *Assume A1: State-dependent receptions, heterogeneous receivers. Given a fixed number of M packets, the expected total delay to broadcast those packets is minimized over all block partitions $(m, \mathbf{c})$ by using a single block of M packets.*

Proof. Following the notation in previous proofs, let $\mathbb{E}[Y_{n:n}^{(c_i)}]$ be the expected time to broadcast a block of c_i packets under RLC. By the monotonicity propositions:

$$\sum_{i=1}^{m} \mathbb{E}[Y_{n:n}^{(c_i)}] = \sum_{i=1}^{m} \frac{\mathbb{E}[Y_{n:n}^{(c_i)}]}{c_i} c_i \geq \sum_{i=1}^{m} \frac{\mathbb{E}[Y_{n:n}^{(M)}]}{M} c_i = \mathbb{E}[Y_{n:n}^{(M)}], \tag{3.72}$$

which establishes the desired inequailty for all feasible block partitions $(m, \mathbf{c})$. □

Having established $\mathbb{E}[Y_{n:n}^{(c)}]/c > \mathbb{E}[Y_{n:n}^{(c+1)}]/(c+1)$ for all c, it is natural to wonder if in fact a stochastic ordering holds. The following proposition is a partial negative answer for the case of infinite field size.

Proposition 18. *Assume A2: State-independent receptions, heterogeneous receivers. The RVs $Y_j^{(c)}/c$ and $Y_j^{(c+1)}/(c+1)$ are not stochastically ordered for any c and j.*

Proof. We use the shorthand notation $\hat{Y}_j^{(c)} \equiv Y_j^{(c)}/c$. We first show a stochastic ordering equivalence among RVs $\hat{Y}_j^{(c)}$, $\hat{Y}_j^{(c+1)}$, and X_j:

$$\left(\hat{Y}_j^{(c)} \geqslant_{\text{st}} \hat{Y}_j^{(c+1)}\right) \Leftrightarrow \left(\hat{Y}_j^{(c)} \geqslant_{\text{st}} X_j\right), \; j \in [n]. \tag{3.73}$$

We prove the equivalence $\leq_{\text{st}}$; the proof for $\geq_{\text{st}}$ is similar. Decompose $\hat{Y}_j^{(c+1)}$ as

$$\hat{Y}_j^{(c+1)} = \frac{1}{c+1}\left(X_{j,1}^{(c+1)} + \sum_{k=2}^{c+1} X_{j,k}^{(c+1)}\right) \overset{\text{d}}{=} \frac{1}{c+1}\left(X_j + \sum_{k=1}^{c} X_{j,k}^{(c)}\right) = \frac{1}{c+1}X_j + \frac{c}{c+1}\hat{Y}_j^{(c)}, \quad (3.74)$$

where the equality in distribution holds due to the assumption that the field size is infinite. Next, decompose $\hat{Y}_j^{(c)} = \frac{1}{c+1}\hat{Y}_j^{(c)} + \frac{c}{c+1}\hat{Y}_j^{(c)}$. Suppose $\hat{Y}_j^{(c)} \leq_{\text{st}} \hat{Y}_j^{(c+1)}$. Substitution of these two decompositions under the assumed stochastic ordering yields

$$\frac{1}{c+1}\hat{Y}_j^{(c)} + \frac{c}{c+1}\hat{Y}_j^{(c)} \leq_{\text{st}} \frac{1}{c+1}X_j + \frac{c}{c+1}\hat{Y}_j^{(c)}, \quad (3.75)$$

which is clearly equivalent to $\hat{Y}_j^{(c)} \leq_{\text{st}} X_j$. Next, suppose $\hat{Y}_j^{(c)} \leq_{\text{st}} X_j$. Then reversing the equivalences used in the proof of the forward direction yields $\hat{Y}_j^{(c)} \leq_{\text{st}} \hat{Y}_j^{(c+1)}$.

We now show how (3.73) leads to a contradiction. Due to Prop. 16, it has to hold that $\hat{Y}_j^{(c)} \geq_{\text{st}} \hat{Y}_j^{(c+1)}$ (note by definition $\Delta^{(c)} = \mathbb{E}\left[\hat{Y}_{n:n}^{(c)}\right] - \mathbb{E}\left[\hat{Y}_{n:n}^{(c+1)}\right]$ and recall properties of stochastic ordering discussed in §3.2.2). According to (3.73) we have $\hat{Y}_j^{(c)} \geq_{\text{st}} X_j \; \forall j$, which gives $\mathbb{E}\left[\hat{Y}_{n:n}^{(c)}\right] - \mathbb{E}\left[X_{n:n}\right] \geq 0$. But this is a contradiction by repeated application of Prop. 16 (also recall Prop. 15). Therefore we conclude that $Y_j^{(c)}/c$ and $Y_j^{(c+1)}/(c+1)$ are not stochastically ordered for any c and j. □

Remark 22. *Our proof techniques for both Prop. 16 and 17 are based on the inequality that the maximum (over row indices $j \in [n]$) of the sum of the first c entries, $\max_{j\in[n]}(X_{j,1} + \cdots + X_{j,c})$, is lower bounded by $X_{\hat{J},1} + \cdots + X_{\hat{J},c}$, where $\hat{J}$ is a (random) row index that maximizes the sum of the first $c+1$ entries. As illustrated in the proofs of these propositions, application of this inequality requires careful manipulation of the distribution of this random index $\hat{J}$. From the definition of $\Delta^{(c)}$, one might be led to hope that a simpler inequality may suffice to establish $\Delta^{(c)} > 0$, possibly leading to a simpler proof. The purpose of this remark is to show that at least one such simpler inequality is insufficient. Suppose the field size is infinite. Applying the decomposition in* (3.74) *to the definition*

of $\Delta^{(c)}$ *gives*

$$\begin{aligned}\Delta^{(c)} &= \mathbb{E}\left[\frac{1}{c}Y_{n:n}^{(c)} - \frac{1}{c+1}\max_{j\in[n]}\left(X_j + Y_j^{(c)}\right)\right] \\ &\geq \mathbb{E}\left[\frac{1}{c}Y_{n:n}^{(c)} - \frac{1}{c+1}\max_{j\in[n]}X_j - \frac{1}{c+1}\max_{j\in[n]}Y_j^{(c)}\right] \quad (3.76) \\ &= \frac{1}{c+1}\mathbb{E}\left[\frac{1}{c}Y_{n:n}^{(c)} - X_{n:n}\right] \quad (3.77)\end{aligned}$$

where the inequality used in (3.76) *is* $\max_j(a_j + b_j) \leq \max_j a_j + \max_j b_j$. *But, repeated application of Prop. 16 shows* (3.77) *is negative, and therefore the inequality* (3.76) *is too weak to establish* $\Delta^{(c)} > 0$.

We now turn our focus to the sample path relationship of these quantities. Consider the total delay to transmit m packets using blocklengths c, c', where without loss of generality we suppose $c' > c$. Using a blocklength c (c') requires $\lceil m/c \rceil$ ($\lceil m/c' \rceil$) blocks, respectively, with the last block possibly being a partial block. We consider two cases: *i)* $m \leq c'$ (Lemma 1) and *ii)* $m > c'$ (Prop. 19), and assume the field size d to be infinite for both. Under this assumption the only randomness in the delay is from the erasure or nonerasure of each transmission to each receiver, and does not include the randomness from the random linear combinations. We introduce some notation. Let $T_{n:n}^{(c,m)}(\omega)$, $T_{n:n}^{(c',m)}(\omega)$ be the delay to broadcast a workload of m packets using a blocklength c, c', respectively, for realization ω.

Lemma 1. *Assume A2: State-independent receptions, heterogeneous receivers. Consider any sample path (realization)* ω *of erasures and nonerasures to each receiver over the sequence of transmissions. The total time to complete the transmission of a workload of* m *packets under blocklengths* c, c' *with* $c < c'$ *and* $m \leq c'$ *obeys* $T_{n:n}^{(c',m)}(\omega) \leq T_{n:n}^{(c,m)}(\omega)$ *for all realizations* $\omega \in \Omega$.

Proof. We consider two cases: *i)* $m \leq c$, and *ii)* $c < m \leq c'$. In the first case, the increase in the blocklength from c to c' will have no effect on the completion time of the workload, since a single block suffices for both blocklengths, thus $T_{n:n}^{(c',m)}(\omega) \leq T_{n:n}^{(c,m)}(\omega)$ for all $\omega \in \Omega$. In the second case, $c < m \leq c'$ means under blocklength c there exists a positive integer k and a nonnegative integer l such that $m = kc + l$, i.e., $k + 1$ (or k, if $l = 0$) blocks are required, with the first k blocks having

length c and the last $(k+1)^{\text{th}}$ block having length l. At most one block is required to handle the workload of m packets under blocklength c'. Let $y^{(c)}_{n:n,i}$ for $i \in [k]$ and $y^{(l)}_{n:n,k+1}$ be the durations of the $(k+1)^{\text{th}}$ block under blocklength c, and

$$t^{(c,m)}_{n:n} = T^{(c,m)}_{n:n}(\omega) = \sum_{i=1}^{k} y^{(c)}_{n:n,i} + y^{(l)}_{n:n,k+1} \tag{3.78}$$

the time at which the workload is completed under blocklength c. Analogously, $t^{(c',m)}_{n:n} = T^{(c',m)}_{n:n}(\omega) = y^{(c')}_{n:n,1}$, is the completion time under c'. We must show $t^{(c',m)}_{n:n} \leq t^{(c,m)}_{n:n}$ for all ω. Let $r_j(t)$ be the number of receptions by receiver j by time t and observe $r_j(t^{(c,m)}_{n:n}) \geq m$ for all $j \in [n]$. Let $t^{(c',m)}_j = \min\{t : r_j(t) = m\}$, and thus $t^{(c',m)}_j \leq t^{(c,m)}_{n:n}$, and $t^{(c',m)}_{n:n} = \max_{j\in[n]} t^{(c',m)}_j \leq t^{(c,m)}_{n:n}$. □

Proposition 19. *Assume A2: State-independent receptions, heterogeneous receivers. Consider any sample path (realization) ω of erasures and nonerasures to each receiver over the sequence of transmissions. The total time to complete the transmission of a workload of m packets under blocklengths c, c' with $c < c' < m$ obeys $T^{(c',m)}_{n:n}(\omega) \leq T^{(c,m)}_{n:n}(\omega)$ for all realizations $\omega \in \Omega$ iff $c' = kc$ for some integer $k \geq 2$.*

The proof of Prop. 19 is found in the Appendix.

Remark 23. *It follows that the total delay under RLC (for any blocklength $c' \geq 2$) is no worse than that of UT for all sample paths, provided $d = \infty$.*

In summary, Prop. 16 and 17 establish that the expected delay per packet is nonincreasing in the blocklength c, Prop. 18 establishes the random delay per packet sequence $\{Y^{(c)}_j/c\}_c$ is not stochastically ordered, while Prop. 19 establishes that the delay per packet is not necessarily nonincreasing in c on a sample path basis. Instead, sample path ordering of delay per packet is only guaranteed when the blocklength is increased by some integer multiple.

3.5.3 Bounds on expected delay per packet

In the previous subsection we established that the expected delay per packet, $\mathbb{E}[Y^{(c)}_{n:n}]/c$, is decreasing in the blocklength c for any finite n. In this subsection we supply lower and upper bounds on

$\mathbb{E}[Y_{n:n}^{(c)}]/c$ that provide a more explicit characterization of the dependence of the delay per packet on the blocklength. These bounds will be shown to be (almost) asymptotically tight in c, in that the asymptotic difference between the lower and upper bounds is one.

In this subsection we restrict our attention to Assumption A3 (state-independent receptions, homogeneous receivers). Recall in the Prop. 1 we established the stochastic ordering $\frac{1}{\phi(q)}\tilde{Y}^{(c)} \leq_{\text{st}} Y^{(c)} \leq_{\text{st}} \frac{1}{\phi(q)}\tilde{Y}^{(c)} + c$. When $d = \infty$ the RVs $Y^{(c)}, \tilde{Y}^{(c)}$ are NegBin(c, q) and Gamma$(c, 1)$, respectively. It follows

$$\frac{1}{\phi(q)}\frac{\mathbb{E}[\tilde{Y}_{n:n}^{(c)}]}{c} \leq \frac{\mathbb{E}[Y_{n:n}^{(c)}]}{c} \leq \frac{1}{\phi(q)}\frac{\mathbb{E}[\tilde{Y}_{n:n}^{(c)}]}{c} + 1, \tag{3.79}$$

Specializing Prop. 13 to the homogeneous case yields $\lim_{c\to\infty} \mathbb{E}[Y_{n:n}^{(c)}]/c = 1/q$ and $\lim_{c\to\infty} \mathbb{E}[\tilde{Y}_{n:n}^{(c)}]/c = 1$, which gives the asymptotic ordering

$$\lim_{c\to\infty} \frac{1}{\phi(q)}\frac{\mathbb{E}[\tilde{Y}_{n:n}^{(c)}]}{c} = \frac{1}{\phi(q)} \leq \lim_{c\to\infty} \frac{\mathbb{E}[Y_{n:n}^{(c)}]}{c} = \frac{1}{q} \leq \lim_{c\to\infty} \frac{1}{\phi(q)}\frac{\mathbb{E}[\tilde{Y}_{n:n}^{(c)}]}{c} + 1 = \frac{1}{\phi(q)} + 1. \tag{3.80}$$

The fact that $\frac{1}{\phi(q)} \leq \frac{1}{q} \leq \frac{1}{\phi(q)} + 1,\ \forall q \in (0, 1)$ may also be verified directly. The next proposition gives lower and upper bounds on $\mathbb{E}[\tilde{Y}_{n:n}^{(c)}]/c$ that are both converging to 1.

Proposition 20. *Assume A3: State-independent receptions, homogeneous receivers. For any positive integers n, c we have bounds on the expected delay per packet*

$$\frac{1}{\phi(q)}\tilde{l}(n, c) \leq \frac{1}{c}\mathbb{E}\left[Y_{n:n}^{(c)}\right] \leq \frac{1}{\phi(q)} nQ\left(c + 1, Q^{-1}\left(c, \frac{1}{n}\right)\right) + 1 \leq \frac{1}{\phi(q)}\tilde{u}(n, c) + 1, \tag{3.81}$$

where $\tilde{l}(n, c) \equiv 1 + \sqrt{\frac{\log n}{c}}\left(1 - \left(1 - Q\left(c, c + \sqrt{c \log n}\right)\right)^{n-1}\right)$, and $\tilde{u}(n, c) \equiv 1 + n/\sqrt{2\pi c}$. Those bounds are asymptotically (in c) tight in the sense that:

$$\lim_{c\to\infty} \tilde{l}(n, c) = \lim_{c\to\infty} \tilde{u}(n, c) = 1. \tag{3.82}$$

Proof. We first show the upper bound $\tilde{u}(n, c)$ is tight. Specializing $r = 1$ in the upper bound in

Prop. 8, we have

$$\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] \leq s + n\left((c-s)Q(c,s) + \frac{s^c \mathrm{e}^{-s}}{\Gamma(c)}\right), \quad s \geq 0. \tag{3.83}$$

When $r = 1$ the optimal Ross's bound is given by $c + nc\frac{(s^*)^c \mathrm{e}^{-s^*}}{\Gamma(c+1)}$ where $s^* = Q^{-1}\left(c, \frac{1}{n}\right)$ is such that $nQ(c, s^*) = 1$. Yet it suffices to show for another choice of s, Ross's bound is tight. For this we need a lower bound on $c!$[22]:

$$\sqrt{2\pi c}\left(\frac{c}{\mathrm{e}}\right)^c \leq c! \leq \frac{1}{\sqrt{1-1/c}}\sqrt{2\pi c}\left(\frac{c}{\mathrm{e}}\right)^c \tag{3.84}$$

Then, dividing both sides of the Ross's upper bound by c and applying the lower bound on the factorial gives

$$\begin{aligned}
\frac{1}{c}\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] &\leq \frac{s}{c} + n\left(\left(1-\frac{s}{c}\right)Q(c,s) + \frac{s^c}{\mathrm{e}^s}\frac{1}{c!}\right) \\
&\leq \frac{s}{c} + n\left(\left(1-\frac{s}{c}\right)Q(c,s) + \frac{s^c}{\mathrm{e}^s}\frac{1}{\sqrt{2\pi c}}\left(\frac{\mathrm{e}}{c}\right)^c\right) \\
&= \frac{s}{c} + n\left(\left(1-\frac{s}{c}\right)Q(c,s) + \frac{1}{\sqrt{2\pi c}}\left(\frac{s}{c}\right)^c \mathrm{e}^{c\left(1-\frac{s}{c}\right)}\right).
\end{aligned} \tag{3.85}$$

Choosing $s = c$ achieves the desired tightness as $c \to \infty$

$$\frac{1}{c}\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] \leq \tilde{u}(n,c) \equiv 1 + n/\sqrt{2\pi c} \to 1. \tag{3.86}$$

Next, we show the lower bound $\tilde{l}(n,c)$ is tight. Specializing $r = 1$ in the lower bound in Prop. 8, we have:

$$\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] \geq t - (t-c)(1-Q(c,t))^{n-1}, \quad t \geq c \tag{3.87}$$

Division by c and reparameterization of t/c by t, and then by $t+1$ gives the de la Cal's bound as

$$\begin{aligned}
\frac{1}{c}\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] &\geq \frac{t}{c} - \left(\frac{t}{c} - 1\right)(1-Q(c,t))^{n-1}, \quad \frac{t}{c} \geq 1 \\
\frac{1}{c}\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] &\geq t - (t-1)(1-Q(c,tc))^{n-1}, \quad t \geq 1 \\
\frac{1}{c}\mathbb{E}[\tilde{Y}^{(c)}_{n:n}] &\geq 1 + t(1-(1-Q(c,c(1+t)))^{n-1}), \quad t \geq 0
\end{aligned} \tag{3.88}$$

Since the quantity $t(1-(1-Q(c,c(1+t)))^{n-1})$ is nonnegative for $t \geq 0$, we argue that the asymptotic tightness of the Ross's bound implies the asymptotic tightness of the de la Cal's bound. To see this, after choosing nonnegative t as a function of c, we write de la Cal's bound as $1+D(c)$ with $D(c) \geq 0$. Observe

$$1 + D(c) \leq \frac{1}{c}\mathbb{E}[\tilde{Y}_{n:n}^{(c)}] \leq \tilde{u}(n,c), \tag{3.89}$$

with $\tilde{u}(n,c) \to 1$ as $c \to \infty$, and hence it must hold $D(c) \to 0$ (the pinch lemma).

Numerical investigation suggests $t(n,c) = c + \sqrt{c \log n}$ can be used in (3.87), which after further normalizing by c becomes:

$$\frac{1}{c}\mathbb{E}[\tilde{Y}_{n:n}^{(c)}] \geq 1 + \sqrt{\frac{\log n}{c}}\left(1 - \left(1 - Q\left(c, c + \sqrt{c \log n}\right)\right)^{n-1}\right) \equiv \tilde{l}(n,c). \tag{3.90}$$

□

Fig. 3.3 shows the (exact) expected delay per packet and the upper and lower bounds from Prop. 20 vs. the blocklength c. Recall the asymptotic delay per packet is $1/q$, which equals 12 (left) and 122 (right), respectively. The lower bound appears to reach the asymptotic value too quickly, while the upper bound appears to track the actual value much better.

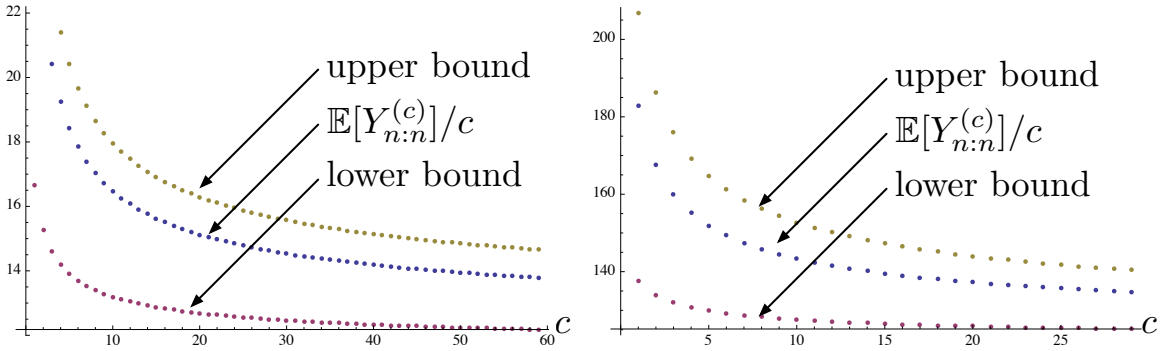

Figure 3.3: The (exact) expected delay per packet and the upper and lower bounds from Prop. 20 vs. the blocklength c for $n = 5$ and $q = 1/12$ (left), and $n = 2$ and $q = 1/122$ (right)

3.6 Dependence of RLC delay on the number of receivers n

The previous section investigated the dependence of RLC delay on the blocklength c, holding the number of receivers n fixed. In this section we investigate the dependence on n, the number of receivers, holding the blocklength c fixed. Throughout this section we make Assumption A3 (state-independent receptions, homogeneous receivers). The key analytical tool in this section is extreme value theory (EVT)[48;49], a closely-related field of order statistics[50], which studies the convergence of minima and maxima of collections of random variables. A difficulty in applying EVT to our framework is the fact that for many common discrete distributions (including geometric, Poisson, negative binomial, etc.) there does not exist a linear normalization such that the normalized maximum order statistic converges in distribution (e.g.,[49], Thm. 1.7.13). This difficulty is circumvented through the use of the stochastic ordering relationship between our discrete delay RV and a continuous analog, as leveraged in §3.4.

As mentioned above, we will again use the stochastic ordering $\frac{1}{\phi(q)}\tilde{Y}^{(c)}_{n:n} \leq_{\text{st}} Y^{(c)}_{n:n} \leq_{\text{st}} \frac{1}{\phi(q)}\tilde{Y}^{(c)}_{n:n} + c$ (Prop. 1), relating the discrete $Y^{(c)}_{n:n}$ to its continuous Gamma-distributed analog $\tilde{Y}^{(c)}_{n:n}$. Thus, the focus in much of this section is the application of EVT (scaling $n \to \infty$) to the continuous RV $\tilde{Y}^{(c)}_{n:n} = \max(\tilde{Y}^{(c)}_1, \ldots, \tilde{Y}^{(c)}_n)$, with iid $\tilde{Y}^{(c)}_j \sim \text{Gamma}(c, 1)$. The EVT framework requires identification of sequences (a_n, b_n) such that the linearly normalized sequence $(\tilde{Y}^{(c)}_{n:n} - b_n)/a_n$ converges in distribution in n to one of three possible extreme value distributions (Gumbel, Fréchet, or Reversed Weibull).

We note that asymptotic expressions for the first two moments of $Y^{(c)}_{n:n}$ on n are derived in[1], with proof techniques adapted from[46], which relies on tools from complex analysis. Our contribution is two-fold. First, we provide an alternate proof technique to that of[1;46], i.e., EVT applied to the continuous distribution $\tilde{Y}^{(c)}_{n:n}$ stochastically ordered with $Y^{(c)}_{n:n}$, which can be used to demonstrate the same dependence upon n of the (first) moment of $Y^{(c)}_{n:n}$. Although the EVT framework naturally gives convergence in distribution of the normalized sequence of RVs, under mild technical conditions, this convergence also implies convergence in r^{th} mean. Using this relationship, we are able to derive asymptotic bounds for the first moment of $Y^{(c)}_{n:n}$ and show that they match reasonably well with those obtained in[1]. Second, we establish that the lower and upper bounds on $\mathbb{E}[\tilde{Y}^{(c)}_{n:n}]$ from §3.4

are asymptotically tight in n as $n \to \infty$. Proving this requires selecting the free parameters (i.e., $s = s_n$ and $t = t_n$) such that the limits can be established. Our choice of (s_n, t_n) is informed by the normalizing sequences (a_n, b_n) identified in the EVT analysis.

3.6.1 Asymptotic delay as $n \to \infty$

The next two propositions are well-known: the first gives the normalizing sequence (a_n, b_n) for the Gamma distribution to be attracted to the standard Gumbel distribution, and the second relates this convergence to convergence of the normalized moments. The subsequent corollary follows immediately from these two propositions and the stochastic ordering between $Y^{(c)}_{n:n}, \tilde{Y}^{(c)}_{n:n}$.

Proposition 21 ([48] §1.5 Example 3)**.** *Let $\tilde{Y}^{(c)}_{n:n}$ be the maximum of n iid $(\tilde{Y}^{(c)}_1, \ldots, \tilde{Y}^{(c)}_n)$ Gamma RVs, with $\tilde{Y}^{(c)}_j \sim \text{Gamma}(c, 1)$. Then*

$$\lim_{n\to\infty} F_{\tilde{Y}^{(c)}_{n:n}}(a_n \tilde{y} + b_n) = \Lambda(\tilde{y}) \equiv \mathrm{e}^{-\mathrm{e}^{-\tilde{y}}}, \ \tilde{y} \in \mathbb{R}, \tag{3.91}$$

for normalizing sequences (a_n, b_n) with

$$a_n = 1, \quad b_n = \log n - \log \Gamma(c) + (c-1) \log \log n. \tag{3.92}$$

In other words, with the above choice of (a_n, b_n), the normalized maximum order statistic $\left(\tilde{Y}^{(c)}_{n:n} - b_n\right)/a_n$ converges in distribution to the standard Gumbel, with CDF $\Lambda(y)$. A random variable $\tilde{Z}$ with distribution $F_{\tilde{Z}}$ is said to belong to the domain of attraction of an extreme value distribution (i.e., Gumbel, Fréchet, or Reversed Weibull) if there is a linear scaling (a_n, b_n) such that $(\tilde{Z}_{n:n} - b_n)/a_n$ converges in distribution to that extreme value distribution, where $\tilde{Z}_{n:n} = \max(\tilde{Z}_1, \ldots, \tilde{Z}_n)$. The next proposition states that, subject to a mild technical condition, if $\tilde{Z} \sim F_{\tilde{Z}}$ belongs to the Gumbel domain of attraction, then its normalized moments converge to a moment-specific constant.

Proposition 22 ([48] Prop. 2.1)**.** *If $\tilde{Z} \sim F_{\tilde{Z}}$ belongs to the Gumbel domain of attraction under scaling*

(a_n, b_n), and if $\int_{-\infty}^{0} |z|^r \mathrm{d}F_{\tilde{Z}}(z) < \infty$ for some $r \in \mathbb{N}$, then

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{\tilde{Z}_{n:n} - b_n}{a_n}\right)^r\right] = (-1)^r \Gamma^{(r)}(1), \tag{3.93}$$

where $\tilde{Z}_{n:n} = \max(\tilde{Z}_1, \ldots, \tilde{Z}_n)$, and $\Gamma^{(r)}(1)$ is the r^{th} derivative of the gamma function evaluated at 1.

Note in particular $(-1)^1\Gamma^{(1)}(1) = \gamma$ for $\gamma \approx 0.5772$ the Euler-Mascheroni constant, and $(-1)^2\Gamma^{(2)}(1) = \gamma^2 + \pi^2/6 \approx 1.9781$. Since $\tilde{Y}_j^{(c)}$ is a nonnegative RV, the technical condition is satisfied.

Corollary 5. *Assume A3: State-independent receptions, homogeneous receivers. The following lower and upper bounds hold for the asymptotic in n scaled r^{th} power (for r odd) of $Y_{n:n}^{(c)} = \max(Y_1^{(c)}, \ldots, Y_n^{(c)})$, for iid $(Y_1^{(c)}, \ldots, Y_n^{(c)})$ with $Y_j^{(c)} \sim \mathrm{NegBin}(c, q)$:*

$$(-1)^r\Gamma^{(r)}(1) \leq \lim_{n\to\infty} \mathbb{E}\left[\left(\phi(q)Y_{n:n}^{(c)} - b_n\right)^r\right] \leq \sum_{s=0}^{r} \binom{r}{s} (-1)^s \Gamma^{(s)}(1) (\phi(q)c)^{r-s}, \tag{3.94}$$

for b_n in (3.92) *and $\Gamma^{(r)}(1)$ the r^{th} derivative of the gamma function evaluated at* 1.

Proof. Multiplying the stochastic ordering $\frac{1}{\phi(q)}\tilde{Y}_{n:n}^{(c)} \leq_{\text{st}} Y_{n:n}^{(c)} \leq_{\text{st}} \frac{1}{\phi(q)}\tilde{Y}_{n:n}^{(c)} + c$ by $\phi(q)$, subtracting b_n, raising to the r^{th} power, and applying the binomial theorem to the upper bound gives

$$\left(\tilde{Y}_{n:n}^{(c)} - b_n\right)^r \leq_{\text{st}} \left(\phi(q)Y_{n:n}^{(c)} - b_n\right)^r \leq_{\text{st}} \left(\left(\tilde{Y}_{n:n}^{(c)} - b_n\right) + \phi(q)c\right)^r = \sum_{s=0}^{r} \binom{r}{s} \left(\tilde{Y}_{n:n}^{(c)} - b_n\right)^s (\phi(q)c)^{r-s} \tag{3.95}$$

Taking expectations preserves stochastic order, and we apply linearity of expectation to the upper bound:

$$\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)} - b_n\right)^r\right] \leq \mathbb{E}\left[\left(\phi(q)Y_{n:n}^{(c)} - b_n\right)^r\right] \leq \sum_{s=0}^{r} \binom{r}{s} \mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)} - b_n\right)^s\right] (\phi(q)c)^{r-s}. \tag{3.96}$$

Taking limits and applying Prop. 22 gives the corollary. □

The corollary is only established for r odd on account of the fact that the function z^r is increasing in its argument for all $z \in \mathbb{R}$ only for r odd, i.e., it is decreasing in z for $z < 0$ for r even. For $r = 1$,

Corollary 5 gives

$$0 \leq \lim_{n\to\infty} \left(\mathbb{E}[Y_{n:n}^{(c)}] - \frac{b_n + \gamma}{\phi(q)} \right) \leq c, \tag{3.97}$$

which captures the same dependence upon n as given in the expression for m_1^{RBC} (i.e., $\mathbb{E}[Y_{n:n}^{(c)}]$) in Proposition 4 of[1]. To see this, recall

$$m_1^{\text{RBC}} = \text{lq}\,(T) + \frac{1}{2} + \frac{\gamma}{\phi(q)} + h\,(\text{lq}\,(T)) + o\,(1)\,, \tag{3.98}$$

where $\text{lq}(\cdot) = \log_{\frac{1}{1-q}}(\cdot)$,

$$T = n \left(\frac{q}{1-q} \right)^{c-1} \left(\log_{\frac{1}{1-q}} n \right)^{c-1} (c-1)!^{-1}, \tag{3.99}$$

and h is a periodic C^∞-function of period 1 and mean value 0, whose Fourier coefficients are $\hat{h}(k) = \frac{1}{\log(1-q)} \Gamma\left(\frac{2\mathrm{i}k\pi}{\log(1-q)} \right)$. Observe $\text{lq}\,(T)$ can be rewritten as:

$$\begin{aligned} \text{lq}\,(T) &= \frac{1}{\phi(q)} \log n + \frac{1}{\phi(q)} (c-1) (\log q + \phi(q)) + \frac{1}{\phi(q)} (c-1) \log\log n - \frac{1}{\phi(q)} (c-1) \log \phi(q) - \frac{1}{\phi(q)} \log (c-1)! \\ &= \frac{1}{\phi(q)} \left(b_n + (c-1) \log \left(\frac{q}{(1-q)\phi(q)} \right) \right). \end{aligned} \tag{3.100}$$

We now give a lemma which *i*) implies the scaling of $\mathbb{E}\left[\left(Y_{n:n}^{(c)} \right)^r \right]$ w.r.t. n is $\left(\frac{1}{\phi(q)} \log n \right)^r$ (Prop. 23), and *ii*) is central to proving the subsequent two propositions showing the asymptotic tightness as $n \to \infty$ of the lower and upper bounds on the r^{th} moment of $\tilde{Y}_{n:n}^{(c)}$.

Lemma 2. *Suppose a sequence of (discrete or continuous) random variables (Z_n) is such that*

$$\lim_{n\to\infty} \mathbb{E}\left[\left(\frac{Z_n - b_n}{a_n} \right)^r \right] = c_r \tag{3.101}$$

for sequences (a_n, b_n) independent of r and (c_r) independent of n, where $a_n = o(b_n)$. Then

$$\lim_{n\to\infty} \frac{\mathbb{E}[Z_n^r]}{b_n^r} = 1, \ r \in \mathbb{N}. \tag{3.102}$$

Proof. Using (3.101) and $a_n = o(b_n)$ gives

$$\lim_{n\to\infty}\left(\frac{a_n}{b_n}\right)^r\left(\mathbb{E}\left[\left(\frac{Z_n-b_n}{a_n}\right)^r\right]-c_r\right)=\lim_{n\to\infty}\mathbb{E}\left[\left(\frac{Z_n}{b_n}-1\right)^r\right]=0. \tag{3.103}$$

Our proof is by induction in r. The above equation immediately gives the base case for $r = 1$. Suppose that (3.102) (with r replaced by s) is true for all $s = 1, \ldots, r-1$. We show this is sufficient to establish the same is true for $s = r$. Using the last equality in (3.103), the binomial theorem, and the induction hypothesis gives the conclusion:

$$\begin{aligned} 0 &= \lim_{n\to\infty}\mathbb{E}\left[\sum_{s=0}^{r}\binom{r}{s}\left(\frac{Z_n}{b_n}\right)^s(-1)^{r-s}\right]=\lim_{n\to\infty}\mathbb{E}\left[\left(\frac{Z_n}{b_n}\right)^r\right]+\sum_{s=0}^{r-1}\binom{r}{s}\left(\lim_{n\to\infty}\mathbb{E}\left[\left(\frac{Z_n}{b_n}\right)^s\right]\right)(-1)^{r-s} \\ &= \lim_{n\to\infty}\mathbb{E}\left[\left(\frac{Z_n}{b_n}\right)^r\right]+\sum_{s=0}^{r-1}\binom{r}{s}(-1)^{r-s}=\lim_{n\to\infty}\mathbb{E}\left[\left(\frac{Z_n}{b_n}\right)^r\right]-1. \end{aligned} \tag{3.104}$$

□

Proposition 23. *Assume A3: State-independent receptions, homogeneous receivers. As $n \to \infty$, the scaling of the r^{th} moment of RLC delay $Y_{n:n}^{(c)}$ is: $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right] \sim \left(\frac{1}{\phi(q)}\log n\right)^r$.*

Proof. Raising the stochastic ordering $\frac{1}{\phi(q)}\tilde{Y}_{n:n}^{(c)} \leq_{\text{st}} Y_{n:n}^{(c)} \leq_{\text{st}} \frac{1}{\phi(q)}\tilde{Y}_{n:n}^{(c)} + c$ to the r^{th} power, applying binomial theorem and taking expectations gives

$$\frac{1}{\phi(q)^r}\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]\leq\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]\leq\sum_{p=0}^{r}\binom{r}{p}\frac{1}{\phi(q)^p}\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^p\right]c^{r-p}.$$

Now dividing through by b_n^r for b_n in (3.92) and taking the limit as $n \to \infty$, we have on the left side of the inequality: $\frac{1}{\phi(q)^r}\lim_{n\to\infty}\frac{\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]}{b_n^r} = \frac{1}{\phi(q)^r}$ by Lem. 2, and furthermore since $b_n \to \infty$ the only term that survives on the right side of the inequality is the one corresponding to $p = r$. Therefore

$$\lim_{n\to\infty}\frac{\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]}{b_n^r}=\frac{1}{\phi(q)^r}, \tag{3.105}$$

Finally the observation $b_n \sim \log n$ concludes the proof. □

3.6.2 Bounds on moments of delay as $n \to \infty$

Prop. 24 and Prop. 25 establish that the lower and upper bounds on $\mathbb{E}\left[\left(Y_{n:n}^{(c)}\right)^r\right]$, resulting from application of the inequalities in Prop. 26 and Prop. 27 to the Gamma RV stochastic ordering from Prop. 1, can be made (almost) asymptotically tight as $n \to \infty$ for fixed c.

Proposition 24. *Assume A3: State-independent receptions, homogeneous receivers. There exists a parameter sequence (t_n) such that the lower bound on $\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]$ from (3.171) can be made to be asymptotically tight in n for every integer $r \geq 1$.*

Proof. Fix integer $r \geq 1$. The lower bound on $\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]$ may be written

$$\begin{aligned}\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right] &\geq t - \left(t - \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]\right)\mathbb{P}\left(\left(\tilde{Y}_{n-1:n-1}^{(c)}\right)^r \leq t\right) \\ &\geq t - \left(t - \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]\right)\mathbb{P}\left(\tilde{Y}_{n-1:n-1}^{(c)} \leq t^{\frac{1}{r}}\right) \end{aligned} \tag{3.106}$$

for any $t \geq \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]$. We establish the asymptotic tightness of this lower bound as $n \to \infty$ by showing the existence of $t_n \to \infty$ such that

$$\lim_{n\to\infty} d(n, t_n) \equiv \lim_{n\to\infty} \frac{t_n - \left(t_n - \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]\right)\mathbb{P}\left(\tilde{Y}_{n-1:n-1}^{(c)} \leq t_n^{\frac{1}{r}}\right)}{\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]} = 1. \tag{3.107}$$

In particular, fix $t_n = (a_{n-1}\tilde{y} + b_{n-1})^r = (\tilde{y} + b_{n-1})^r$ for (a_n, b_n) in (3.92), and $\tilde{y} \in \mathbb{R}$ a constant to be chosen later. Note that for any $\tilde{y}$ there exists an $N_r \in \mathbb{N}$ such that $t_n \geq \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]$ for all $n \geq N_r$, and thus this choice of t_n is asymptotically feasible. Under this scaling we may apply Prop. 21:

$$\lim_{n\to\infty} \mathbb{P}\left(\tilde{Y}_{n-1:n-1}^{(c)} \leq t_n^{\frac{1}{r}}\right) = \Lambda(\tilde{y}). \tag{3.108}$$

Next, by Lemma 2,

$$\lim_{n\to\infty} \frac{\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]}{b_n^r} = 1. \tag{3.109}$$

Substitution of the above two limits into (3.107) and rearranging gives

$$\begin{aligned}\lim_{n\to\infty} d(n,t_n) &= \lim_{n\to\infty}\left[(1-\Lambda(\tilde{y}))\left(\frac{\tilde{y}+b_{n-1}}{b_n}\right)^r + \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]\Lambda(\tilde{y})\frac{1}{b_n^r}\right] \\ &= (1-\Lambda(\tilde{y}))\left(\tilde{y}\lim_{n\to\infty} b_n^{-1} + \lim_{n\to\infty}\frac{b_{n-1}}{b_n}\right)^r + \mathbb{E}\left[\left(\tilde{Y}^{(c)}\right)^r\right]\Lambda(\tilde{y})\lim_{n\to\infty}\frac{1}{b_n^r} \\ &= 1-\Lambda(\tilde{y}), \end{aligned} \tag{3.110}$$

where the above limits are justified under the observation that $b_n \to \infty$ as $n \to \infty$, and

$$\lim_{n\to\infty}\frac{b_{n-1}}{b_n} = \lim_{n\to\infty}\frac{-\log\Gamma(c)+\log(n-1)+(c-1)\log\log(n-1)}{-\log\Gamma(c)+\log n+(c-1)\log\log n} = 1. \tag{3.111}$$

Now let $\tilde{y}$ approach $-\infty$ to obtain the desired limit

$$\lim_{\tilde{y}\to-\infty}\lim_{n\to\infty} d(n,t_n) = \lim_{\tilde{y}\to-\infty} 1-\Lambda(\tilde{y}) = 1. \tag{3.112}$$

□

Proposition 25. *Assume A3: State-independent receptions, homogeneous receivers. There exists a parameter sequence* (s_n) *such that the upper bound on* $\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]$ *from* (3.172) *can be made to be asymptotically tight in* n *for every integer* $r \geq 1$*. In particular* $s_n = b_n^r$ *is sufficient to establish the bound is asymptotically tight, as well as asymptotically optimal, in the asymptotic sense of* (3.173)*.*

Proof. Fix integer $r \geq 1$, and set $s_n = b_n^r$ for b_n in (3.92). Our obligation is to show

$$\lim_{n\to\infty}\frac{b_n^r + n\int_{b_n^r}^{\infty}\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r > t\right)\mathrm{d}t}{\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]} = 1. \tag{3.113}$$

First, by Lemma 2

$$\lim_{n\to\infty}\frac{\mathbb{E}\left[\left(\tilde{Y}_{n:n}^{(c)}\right)^r\right]}{b_n^r} = 1. \tag{3.114}$$

Thus, it suffices to show

$$\lim_{n\to\infty}\frac{n}{b_n^r}\int_{b_n^r}^{\infty}\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r>t\right)\mathrm{d}t=\lim_{n\to\infty}\frac{\int_{b_n^r}^{\infty}\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r>t\right)\mathrm{d}t}{\frac{b_n^r}{n}}=0. \tag{3.115}$$

Observe the limit of both numerator and denominator in the above expression are zero, and thus we may apply L'Hopital's rule, and then Leibniz's rule:

$$\lim_{n\to\infty}\frac{n}{b_n^r}\int_{b_n^r}^{\infty}\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r>t\right)\mathrm{d}t=\lim_{n\to\infty}\frac{\frac{\mathrm{d}}{\mathrm{d}n}\int_{b_n^r}^{\infty}\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r>t\right)\mathrm{d}t}{\frac{\mathrm{d}}{\mathrm{d}n}\left(\frac{b_n^r}{n}\right)}=\lim_{n\to\infty}\frac{-\mathbb{P}\left(\tilde{Y}^{(c)}>b_n\right)\frac{\mathrm{d}}{\mathrm{d}n}b_n^r}{\frac{\mathrm{d}}{\mathrm{d}n}\left(\frac{b_n^r}{n}\right)}. \tag{3.116}$$

As will be shown below

$$\lim_{n\to\infty}n\mathbb{P}\left(\tilde{Y}^{(c)}>b_n\right)=1. \tag{3.117}$$

Substituting (3.117) into (3.116) gives

$$\lim_{n\to\infty}\frac{n}{b_n^r}\int_{b_n^r}^{\infty}\mathbb{P}\left(\left(\tilde{Y}^{(c)}\right)^r>t\right)\mathrm{d}t=\lim_{n\to\infty}\frac{-\frac{1}{n}\frac{\mathrm{d}}{\mathrm{d}n}b_n^r}{\frac{\mathrm{d}}{\mathrm{d}n}\left(\frac{b_n^r}{n}\right)}=\lim_{n\to\infty}\frac{-\frac{1}{n}\frac{\mathrm{d}}{\mathrm{d}n}b_n^r}{\frac{1}{n}\frac{\mathrm{d}}{\mathrm{d}n}b_n^r-\frac{b_n^r}{n^2}}=\lim_{n\to\infty}\frac{1}{\frac{b_n^r}{n\frac{\mathrm{d}}{\mathrm{d}n}b_n^r}-1}=\lim_{n\to\infty}\frac{1}{\frac{b_n}{rn\frac{\mathrm{d}}{\mathrm{d}n}b_n}-1} \tag{3.118}$$

To establish the desired limit of 0, it suffices to show the first term in the denominator grows to infinity in n:

$$\lim_{n\to\infty}\frac{b_n}{rn\frac{\mathrm{d}}{\mathrm{d}n}b_n}=\frac{1}{r}\lim_{n\to\infty}\frac{-\log\Gamma(c)+\log n+(c-1)\log\log n}{1+\frac{c-1}{\log n}}=\infty. \tag{3.119}$$

It remains to establish (3.117). It is well-known that

$$\lim_{s\to\infty}\frac{Q(c,s)}{\frac{s^{c-1}\mathrm{e}^{-s}}{\Gamma(c)}}=1, \tag{3.120}$$

where $Q(c,s)=\mathbb{P}(\tilde{Y}^{(c)}>s)$. Substituting this into (3.117) gives

$$\lim_{n\to\infty}nQ(c>b_n)=\lim_{n\to\infty}n\frac{b_n^{c-1}\mathrm{e}^{-b_n}}{\Gamma(c)}=\lim_{n\to\infty}\left(\frac{b_n}{\log n}\right)^{c-1}=\lim_{n\to\infty}\left(-\frac{\log\Gamma(c)}{\log n}+1+(c-1)\frac{\log\log n}{\log n}\right)^{c-1}=1. \tag{3.121}$$

We observe that for any given n the optimal s^* satisfies $n\mathbb{P}(\tilde{Y}^{(c)} > s^*) = 1$; it is in this sense that $s_n = b_n^r$ is not only sufficient for the bound to be asymptotically tight, but is also the asymptotically optimal choice for s. □

Fig. 3.4 shows several of the bounds discussed in this section vs. the number of receivers, n. The plots attest to the fact that our upper and lower bounds do enclose the exact expected delay for all n. Note that our lower bound is in fact better than the approximation m_1^{RBC} from [1] for larger c, but not for smaller c.

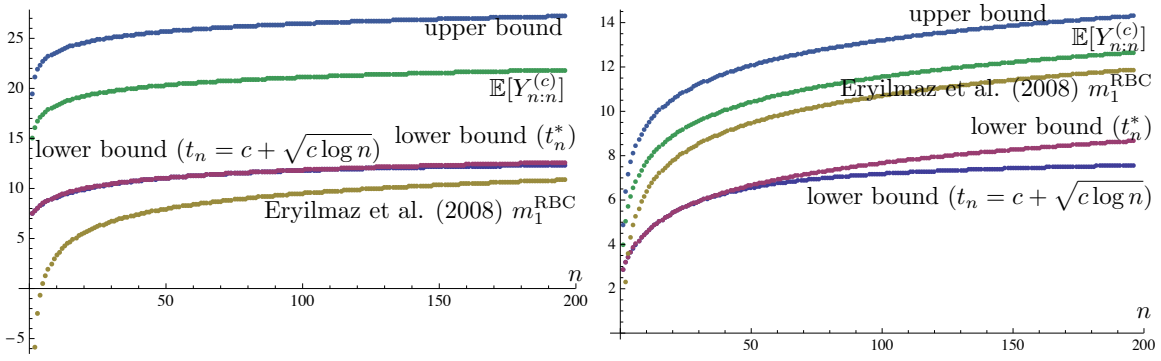

Figure 3.4: The exact block delay ($\mathbb{E}[Y_{n:n}^{(c)}]$), the optimized upper bound, the lower bound with $t_n = c + \sqrt{c \log n}$, the numerically optimized lower bound (with t_n^*), and the approximation m_1^{RBC} from [1], for $c = 12$ and $q = 4/5$ (left) and $c = 2$ and $q = 1/2$ (right).

3.7 Appendix

3.7.1 Properties of regularized incomplete Beta function $I_x(a, b)$

The following definitions apply for $a > 0$, $b > 0$ and $x \in (0, 1)$:

$$I_x(a,b) = \frac{B(x;a,b)}{B(a,b)}, \quad B(x;a,b) = \int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t, \tag{3.122}$$

where $B(x;a,b)$ is the incomplete Beta function and $B(a,b) = B(1;a,b)$ is the Beta function.

Lemma 3. *$I_x(a,b)$ is increasing in b.*

Proof. As

$$\frac{\partial}{\partial b} I_x(a,b) = \frac{\frac{\partial}{\partial b}\left(B(x;a,b)\right) B(a,b) - B(x;a,b)\frac{\partial}{\partial b}\left(B(a,b)\right)}{B^2(a,b)}, \tag{3.123}$$

it suffices to show the positivity of the numerator in the above expression, which becomes:

$$\int_0^x t^{a-1}(1-t)^{b-1}\log(1-t)\mathrm{d}t\cdot\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t-\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t\cdot\int_0^1 t^{a-1}(1-t)^{b-1}\log(1-t)\mathrm{d}t. \tag{3.124}$$

Observe that due to the negativity of $\log(1-t)$ for $t\in(0,1)$, the signs of the above four integrals are respectively: $-,+,+,-$, and hence we need to show:

$$f(x;a,b)\equiv\frac{\int_0^x t^{a-1}(1-t)^{b-1}\log(1-t)\mathrm{d}t}{\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t}>\frac{\int_0^1 t^{a-1}(1-t)^{b-1}\log(1-t)\mathrm{d}t}{\int_0^1 t^{a-1}(1-t)^{b-1}\mathrm{d}t}=f(1;a,b). \tag{3.125}$$

It suffices to show $f(x;a,b)$ is decreasing in x for $x\in(0,1)$, i.e., $\frac{\partial}{\partial x}f(x;a,b)<0$. Application of Leibniz's rule to differentiate $f(x;a,b)$ w.r.t. x gives

$$\frac{\partial}{\partial x}f(x;a,b)=\frac{x^{a-1}(1-x)^{b-1}\log(1-x)\cdot\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t-x^{a-1}(1-x)^{b-1}\int_0^x t^{a-1}(1-t)^{b-1}\log(1-t)\mathrm{d}t}{B^2(x;a,b)}. \tag{3.126}$$

When $x\in(0,1), a>0, b>0$, this partial derivative is negative if

$$\int_0^x t^{a-1}(1-t)^{b-1}\log(1-x)\mathrm{d}t<\int_0^x t^{a-1}(1-t)^{b-1}\log(1-t)\mathrm{d}t. \tag{3.127}$$

It can be seen that $0\le\log(1-x)\le\log(1-t)$ when $0\le t\le x<1$, which means the above inequality holds. Hence we have shown $I_x(a,b)$ is increasing in b when $a>0, b>0$ and $x\in(0,1)$. □

Lemma 4. *The function* $\log I_x(a,b)$ *is concave on* $x\in(0,1)$*, for integers* $a\ge1, b\ge1$.

Proof. We shall verify the second derivative is non-positive. Towards this,

$$\begin{aligned}\frac{\partial}{\partial x}\log I_x(a,b) &= \frac{1}{I_x(a,b)}\frac{\partial}{\partial x}I_x(a,b)\\ \frac{\partial^2}{\partial x^2}\log I_x(a,b) &= \frac{\left(\frac{\partial^2}{\partial x^2}I_x(a,b)\right)I_x(a,b)-\left(\frac{\partial}{\partial x}I_x(a,b)\right)^2}{\left(I_x(a,b)\right)^2},\end{aligned} \tag{3.128}$$

hence $\frac{\partial^2}{\partial x^2}\log I_x(a,b) \leq 0 \iff \left(\frac{\partial^2}{\partial x^2} I_x(a,b)\right) I_x(a,b) \leq \left(\frac{\partial}{\partial x} I_x(a,b)\right)^2$. As

$$\frac{\partial}{\partial x} I_x(a,b) = \frac{1}{B(a,b)} \frac{\partial}{\partial x} B(x;a,b), \quad \frac{\partial^2}{\partial x^2} I_x(a,b) = \frac{1}{B(a,b)} \frac{\partial^2}{\partial x^2} B(x;a,b), \tag{3.129}$$

we need to show:

$$B(x;a,b)\frac{\partial^2}{\partial x^2} B(x;a,b) \leq \left(\frac{\partial}{\partial x} B(x;a,b)\right)^2. \tag{3.130}$$

Recalling the definition of the incomplete Beta function given in (3.122), we have

$$\begin{aligned} \frac{\partial}{\partial x} B(x;a,b) &= x^{a-1}(1-x)^{b-1} \\ \frac{\partial^2}{\partial x^2} B(x;a,b) &= (a-1)x^{a-2}(1-x)^{b-1} - (b-1)x^{a-1}(1-x)^{b-2}. \end{aligned} \tag{3.131}$$

After canceling $x^{a-2}(1-x)^{b-2}$, (3.130) becomes

$$((a-1)(1-x) - (b-1)x) \int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t \leq x^a(1-x)^b. \tag{3.132}$$

Define $x^* = \frac{a-1}{a-1+b-1}$ to satisfy $(a-1)(1-x^*) - (b-1)x^* = 0$ so that $(a-1)(1-x) - (b-1)x \gtrless 0$ iff $x \lessgtr x^*$ (recall $a \geq 1, b \geq 1$). When $x \geq x^*$, the above inequality holds since the LHS is nonpositive.

It remains to consider $x < x^*$. When $a = 1$ and/or $b = 1$, the above inequality can be verified to hold for all x. Therefore, from now on assume $a > 1, b > 1$ and $0 < x < x^*$. Note $x^* < 1$ under these assumptions. We seek to show

$$f(x) \equiv ((a-1)(1-x) - (b-1)x) \int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t - x^a(1-x)^b \leq 0. \tag{3.133}$$

Note $f(0) = 0$, thus it suffices to show $f'(x) \leq 0$, where

$$f'(x) = (2-a-b)\int_0^x t^{a-1}(1-t)^{b-1}\mathrm{d}t + x^{a-1}(1-x)^{b-1}(2x-1). \tag{3.134}$$

The assumption $a > 1$ and $b > 1$ ensures the first term in the sum is nonpositive. We again consider

two cases: $a \leq b$ and $a > b$. For $a \leq b$ (equivalently, $x^* \leq 1/2$), any $x < x^*$ will make the second term in (3.134) nonpositive, and thus $f'(x) \leq 0$.

It remains to consider $a > b$ (equivalently $x^* > 1/2$), and in particular $1/2 < x < x^* < 1$. Observe $f'(1/2) < 0$, so it further suffices to show $f''(x) < 0$ for the above regime of interest. After some algebra, one can show that

$$f''(x) = x^{a-2}(1-x)^{b-2}\left[-(a+b)\left(x - \frac{a}{a+b}\right)^2 + 1 - \frac{ab}{a+b}\right], \tag{3.135}$$

where the sign of $f''(x)$ is determined by the expression in the brackets, which is a concave quadratic taking maximum value of $1 - ab/(a+b)$ at $x = a/(a+b)$. For integers a, b satisfying $a > b > 1$, observe $b > 1 + b/a$ and thus $ab > a + b$, and so the value of the quadratic at its maximum is negative, establishing $f''(x) < 0$ for $x \in (1/2, x^*)$. □

3.7.2 Proof of Prop. 14

Assume A1: State-dependent receptions, heterogeneous receivers. In order to establish (3.62) holds for $q_{j,k} = (1 - d^{k-1-c})q_j$ it suffices to establish $a)$ $\lim_{c\to\infty} \frac{1}{c}\sum_{k=1}^{c}(1 - d^{k-1-c})^{-1} = 1$ and $b)$ $\lim_{c\to\infty} \frac{1}{c}\sum_{k=1}^{c}(1 - d^{k-1-c})^{-2} = 1$. The first limit $a)$ establishes the convergence of the means, and the two limits together, $a)$ and $b)$, establish the convergence of the variances. We establish both $a)$ and $b)$ by sandwiching the sum between lower and upper bounds which are integrals of the function being summed. In particular, if $\{f(k)\}_{k=0}^{c}$ is a nondecreasing sequence in k then

$$\int_0^c f(x)\mathrm{d}x \leq \sum_{k=1}^{c} f(k) \leq \int_1^c f(x)\mathrm{d}x + f(c), \tag{3.136}$$

where the lower (upper) bound follows by viewing $\{f(k)\}$ as a Riemann sum of c terms with unit intervals and heights at the right (left) endpoints, respectively. Both $f_a(k) \equiv (1 - d^{k-1-c})^{-1}$ and $f_b(k) \equiv (1 - d^{k-1-c})^{-2}$ are increasing in k, and so (3.136) holds for $f(k) = f_a(k)$ and $f(k) = f_b(k)$. We now proceed to evaluate the integrals. We first find the indefinite integral of $g_a(x) \equiv \int f_a(x)\mathrm{d}x$ and $g_b(x) \equiv \int f_b(x)\mathrm{d}x$ for $d > 1$ and $c \geq 1$. Use the change of variables $z = d^{x-1-c}$ so that

$\mathrm{d}x = \mathrm{d}z/(z \log d)$, yielding:

$$g_a(x(z)) = \frac{1}{\log d} \int \frac{1}{z(1-z)} \mathrm{d}z, \quad g_b(x(z)) = \frac{1}{\log d} \int \frac{1}{z(1-z)^2} \mathrm{d}z \tag{3.137}$$

for $x(z) \equiv c + 1 + \log z / \log d$. Use partial fraction expansion to get

$$g_a(x(z)) = \frac{1}{\log d} \int \left(\frac{1}{z} + \frac{1}{1-z} \right) \mathrm{d}z, \quad g_b(x(z)) = \frac{1}{\log d} \int \left(\frac{1}{z} + \frac{1}{1-z} + \frac{1}{(1-z)^2} \right) \mathrm{d}z \tag{3.138}$$

which integrates to

$$g_a(x(z)) = \frac{\log z - \log(1-z)}{\log d}, \quad g_b(x(z)) = \frac{\log z - \log(1-z) + 1/(1-z)}{\log d}. \tag{3.139}$$

Substituting back $z = d^{x-1-c}$ yields:

$$g_a(x) = \frac{\log d^{x-1-c} - \log(1-d^{x-1-c})}{\log d}, \quad g_b(x) = \frac{\log d^{x-1-c} - \log(1-d^{x-1-c}) + 1/(1-d^{x-1-c})}{\log d}. \tag{3.140}$$

We next integrate from 0 to c and from 1 to c respectively, and simplify:

$$\begin{aligned}
g_a(x)|_0^c &= \frac{\log(d^{c+1}-1) - \log(d-1)}{\log d} \\
g_a(x)|_1^c &= \frac{\log(d^{c}-1) - \log(d-1)}{\log d} \\
g_b(x)|_0^c &= \frac{\log(d^{c+1}-1) - \log(d-1) + \frac{d(d^c-1)}{(d-1)(d^{c+1}-1)}}{\log d} \\
g_b(x)|_1^c &= \frac{\log(d^{c}-1) - \log(d-1) + \frac{d(d^{c-1}-1)}{(d-1)(d^{c}-1)}}{\log d}
\end{aligned} \tag{3.141}$$

We are interested in the limits of these integrals divided by c as $c \to \infty$. Observe the last term in the numerators of the latter two expressions is dominated by the first term as $c \to \infty$ and may be ignored. It is clear both the numerator and denominator diverge, therefore we differentiate both

and apply L'Hopital's rule. This yields

$$\lim_{c\to\infty} \frac{1}{c} g_a(x)\Big|_0^c = \lim_{c\to\infty} \frac{1}{c} g_a(x)\Big|_1^c = \lim_{c\to\infty} \frac{d^{c+1}}{d^{c+1}-1} = 1, \lim_{c\to\infty} \frac{1}{c} g_b(x)\Big|_0^c = \lim_{c\to\infty} \frac{1}{c} g_b(x)\Big|_1^c = \lim_{c\to\infty} \frac{d^c}{d^c-1} = 1. \tag{3.142}$$

3.7.3 Proof of Prop. 16 (monotonicity for infinite field size)

Assume A2: State-independent receptions, heterogeneous receivers. The proof below holds for an arbitrary $n \times c$ matrix of discrete RVs $\mathbf{Z} = (Z_{j,k})$ such that each row $(Z_{j,k}, k \in [c])$ holds entries iid in k, and the rows are independent of one another. We begin with a lemma that holds for a column-invariant row selection rule χ.

Lemma 5. *Let $\mathbf{Z}$ be an $n \times c$ random matrix with entries $Z_{j,k}$ iid in k and independent in j. Let $J = \chi(\mathbf{Z}) \in [n]$ be the index selected by any column invariant row-index selection rule χ. Then the RVs $(Z_{J,k}, k \in [c])$ in row J are identically distributed.*

Proof. Fix column indices $k, k' \in [c]$ and some $z \in \mathcal{Z}$, the support of the RVs comprising $\mathbf{Z}$. Then:

$$\begin{aligned}
\mathbb{P}(Z_{J,k} = z) &= \sum_{j=1}^{n} \mathbb{P}(Z_{J,k} = z, J = j) \\
&= \sum_{j=1}^{n} \mathbb{P}(Z_{j,k} = z, J = j) \\
&= \sum_{j=1}^{n} \mathbb{P}(J = j | Z_{j,k} = z) \mathbb{P}(Z_{j,k} = z) \\
&= \sum_{j=1}^{n} \mathbb{P}(J = j | Z_{j,k'} = z) \mathbb{P}(Z_{j,k'} = z) \\
&= \mathbb{P}(Z_{J,k'} = z),
\end{aligned} \tag{3.143}$$

where we have used the facts that *i)* χ is column invariant, i.e., $\mathbb{P}(J = j | Z_{j,k} = z) = \mathbb{P}(J = j | Z_{j,k'} = z)$, and *ii)* each row of $\mathbf{Z}$ has entries iid in k, i.e., $\mathbb{P}(Z_{j,k} = z) = \mathbb{P}(Z_{j,k'} = z)$. □

Proof. (of Prop. 16) Let $\mathbf{Z}^{(c)}$ and $\mathbf{Z}^{(c+1)}$ be $n \times c$ and $n \times (c+1)$ matrices, respectively. The entries $Z_{j,k}^{(c)}$ in $\mathbf{Z}^{(c)}$ are assumed iid in k and independent in j, as are the entries $Z_{j,k}^{(c+1)}$ in $\mathbf{Z}^{(c+1)}$. Furthermore, for each $j, j' \in [n]$ and each $k, k' \in [c]$, the RVs $Z_{j,k}^{(c)}$ and $Z_{j',k'}^{(c+1)}$ are assumed iid. Let $\mathbf{Z}_{1:c}^{(c+1)}$ be the

$n \times c$ submatrix of $\mathbf{Z}^{(c+1)}$ obtained by removing the entries in column $c+1$. The above assumptions assert $\mathbf{Z}^{(c)} \stackrel{\mathrm{d}}{=} \mathbf{Z}^{(c+1)}_{1:c}$. As equality in distribution is preserved under any measurable function, say f, it follows that $f(\mathbf{Z}^{(c)}) \stackrel{\mathrm{d}}{=} f(\mathbf{Z}^{(c+1)}_{1:c})$ for all such functions. Finally, equality in distribution ensures equality in expectation, thus $\mathbb{E}[f(\mathbf{Z}^{(c)})] = \mathbb{E}[f(\mathbf{Z}^{(c+1)}_{1:c})]$. This will be used in step (b) of (3.144) below.

Let $\hat{\chi}$ be the column invariant row selection rule that selects the minimum index among the row sum maximizing indices, and let $\hat{J}^{(c+1)} = \hat{\chi}(\mathbf{Z}^{(c+1)})$ be the index selected under $\hat{\chi}$ for $\mathbf{Z}^{(c+1)}$. Then:

$$\begin{aligned}
\Delta^{(c)} &\stackrel{(a)}{=} \mathbb{E}\left[\frac{1}{c}\max_{j\in[n]}\sum_{k=1}^{c} Z^{(c)}_{j,k}\right] - \mathbb{E}\left[\frac{1}{c+1}\max_{j\in[n]}\sum_{k=1}^{c+1} Z^{(c+1)}_{j,k}\right] \\
&\stackrel{(b)}{=} \mathbb{E}\left[\frac{1}{c}\max_{j\in[n]}\sum_{k=1}^{c} Z^{(c+1)}_{j,k}\right] - \mathbb{E}\left[\frac{1}{c+1}\max_{j\in[n]}\sum_{k=1}^{c+1} Z^{(c+1)}_{j,k}\right] \\
&\stackrel{(c)}{=} \mathbb{E}\left[\frac{1}{c}\max_{j\in[n]}\sum_{k=1}^{c} Z^{(c+1)}_{j,k}\right] - \mathbb{E}\left[\frac{1}{c+1}\sum_{k=1}^{c+1} Z^{(c+1)}_{\hat{J}^{(c+1)},k}\right] \\
&\stackrel{(d)}{\geq} \mathbb{E}\left[\frac{1}{c}\sum_{k=1}^{c} Z^{(c+1)}_{\hat{J}^{(c+1)},k}\right] - \mathbb{E}\left[\frac{1}{c+1}\sum_{k=1}^{c+1} Z^{(c+1)}_{\hat{J}^{(c+1)},k}\right] \\
&\stackrel{(e)}{=} 0
\end{aligned} \tag{3.144}$$

where (a) follows by linearity of expectation, (b) follows since $\mathbb{E}[f(\mathbf{Z}^{(c)})] = \mathbb{E}[f(\mathbf{Z}^{(c+1)}_{1:c})]$ for any measurable f, (c) is by definition of the index $\hat{J}^{(c+1)}$, (d) follows since the maximum of a set is no smaller than any of its elements, and (e) follows by applying linearity of expectation and Lemma 5. This establishes $\Delta^{(c)} \geq 0$. Lemma 6 below establishes that the inequality in (3.144) is in fact strict, and thus $\Delta^{(c)} > 0$. □

Lemma 6. *Let $\mathbf{Z}^{(c+1)}$ and $\hat{J}^{(c+1)}$ be as in the proof of Prop. 16. Then:*

$$\mathbb{E}\left[\max_{j\in[n]}\sum_{k=1}^{c} Z^{(c+1)}_{j,k}\right] > \mathbb{E}\left[\sum_{k=1}^{c} Z^{(c+1)}_{\hat{J}^{(c+1)},k}\right]. \tag{3.145}$$

Proof. Let J be a random index in $[n]$ with an arbitrary dependence upon $\mathbf{Z}^{(c+1)}$. Observe that for any J

$$\max_{j\in[n]}\sum_{k=1}^{c} Z^{(c+1)}_{j,k} = \max\left(\max_{j\in[n]\setminus\{J\}}\sum_{k=1}^{c} Z^{(c+1)}_{j,k}, \sum_{k=1}^{c} Z^{(c+1)}_{J,k}\right), \tag{3.146}$$

which we henceforth abbreviate as $Z_1 = \max(Z_2, Z_3)$. The lemma is equivalent to the assertion $\mathbb{E}[Z_1] > \mathbb{E}[Z_3]$, which in turn is equivalent to:

$$\sum_{z=c}^{\infty} \mathbb{P}(Z_1 > z) > \sum_{z=c}^{\infty} \mathbb{P}(Z_3 > z). \tag{3.147}$$

Define $\mathbb{N}_{\geq c} = \{c, c+1, \ldots\}$. It will suffice to show $i)$ $\mathbb{P}(Z_1 > z) \geq \mathbb{P}(Z_3 > z)$ for all $z \in \mathbb{N}_{\geq c}$, and $ii)$ there exists $z' \in \mathbb{N}_{\geq c}$ for which $\mathbb{P}(Z_1 > z') > \mathbb{P}(Z_3 > z')$. The first condition is trivially true since $\mathbb{P}(Z_1 > z) = \mathbb{P}(\max(Z_2, Z_3) > z) \geq \mathbb{P}(Z_3 > z)$. It remains to establish the second condition. Fixing $z' \in \mathbb{N}_{\geq c}$ and conditioning on Z_3 yields

$$\begin{aligned} \mathbb{P}(Z_1 > z') &= \sum_{x \leq z'} \mathbb{P}(Z_1 > z' | Z_3 = x)\mathbb{P}(Z_3 = x) + \sum_{x > z'} \mathbb{P}(Z_1 > z' | Z_3 = x)\mathbb{P}(Z_3 = x) \\ \mathbb{P}(Z_3 > z') &= \sum_{x \leq z'} \mathbb{P}(Z_3 > z' | Z_3 = x)\mathbb{P}(Z_3 = x) + \sum_{x > z'} \mathbb{P}(Z_3 > z' | Z_3 = x)\mathbb{P}(Z_3 = x) \end{aligned} \tag{3.148}$$

After simple manipulations condition $ii)$ becomes

$$\mathbb{P}(Z_1 > z') > \mathbb{P}(Z_3 > z') \;\Leftrightarrow\; \sum_{x \leq z'} \mathbb{P}(Z_2 > z' | Z_3 = x)\mathbb{P}(Z_3 = x) > 0. \tag{3.149}$$

To establish (3.149) it suffices to identify a pair (z', x) with $x \leq z'$ such that $i)$ $\mathbb{P}(Z_3 = x) > 0$ and $\mathbb{P}(Z_2 > z' | Z_3 = x) > 0$. In words, x must be a feasible value for the sum of the first c columns under the row selection rule J, and $z' \geq x$ must strictly lower bound a feasible value for the maximum sum of the first c columns over all rows except J, conditioned on $Z_3 = x$. Fix $\mathbf{Z}^{(c+1)}$ so that $J = j$ and $i \neq j$ is such that $Z_{l,1}^{(c+1)} + \cdots + Z_{l,c}^{(c+1)} \leq Z_{i,1}^{(c+1)} + \cdots + Z_{i,c}^{(c+1)}$ for all $l \neq j$. Set $x = Z_{J,1}^{(c+1)} + \cdots + Z_{J,c}^{(c+1)}$ and $z' < Z_{i,1}^{(c+1)} + \cdots + Z_{i,c}^{(c+1)}$. As long as there exists a realization of $\mathbf{Z}^{(c+1)}$ satisfying the condition on J, the condition on i, and $x \leq z'$, then this suffices to establish the lemma. □

3.7.4 Proof of Prop. 17 (monotonicity for finite field size)

Proof. Assume A1: State-dependent receptions, heterogeneous receivers. Note $\Delta^{(c)} > 0 \Leftrightarrow c(c+1)\Delta^{(c)} > 0$. Fix the field size d to be a finite integer ≥ 2, fix the blocklength $c \in \mathbb{N}$, and fix the

number of receivers $n \in \mathbb{N}$. Let $\mathbf{X}^{(c)}, \mathbf{X}^{(c+1)}$ be $n \times c$ and $n \times (c+1)$ random matrices with entries $X_{j,k}^{(c)} \sim \text{Geo}(q_{j,k}^{(c)})$ and $X_{j,k}^{(c+1)} \sim \text{Geo}(q_{j,k}^{(c+1)})$, respectively. Then:

$$c(c+1)\Delta^{(c)} = \mathbb{E}\left[(c+1)\max_{j\in[n]}\sum_{k=1}^{c} X_{j,k}^{(c)} - c\max_{j\in[n]}\sum_{k=1}^{c+1} X_{j,k}^{(c+1)}\right]. \tag{3.150}$$

By the standing model assumptions, the entries $X_{j,k}^{(c)}$ in $\mathbf{X}^{(c)}$ are independent in j and k, as are the entries $X_{j,k}^{(c+1)}$ in $\mathbf{X}^{(c+1)}$. Furthermore, for each $j, j' \in [n]$ and each $k \in [c]$, $k' \in [c+1]$, the RVs $X_{j,k}^{(c)}$ and $X_{j',k'}^{(c+1)}$ are assumed independent. Recall $q_{j,k}^{(c)} = (1 - d^{k-1-c})q_j$. It is immediate that $X_{j,k}^{(c)} \overset{\text{d}}{=} X_{j,k+1}^{(c+1)}$. Let $\mathbf{X}_{2:c+1}^{(c+1)}$ be the submatrix obtained by removing the entries in column 1 of $\mathbf{X}^{(c+1)}$. The above equality in distribution then gives $\mathbf{X}^{(c)} \overset{\text{d}}{=} \mathbf{X}_{2:c+1}^{(c+1)}$. Further, for any measurable function f, $\mathbb{E}[f(\mathbf{X}^{(c)})] = \mathbb{E}[f(\mathbf{X}_{2:c+1}^{(c+1)})]$. It follows that

$$c(c+1)\Delta^{(c)} = \mathbb{E}\left[(c+1)\max_{j\in[n]}\sum_{k=1}^{c} X_{j,k+1}^{(c+1)} - c\max_{j\in[n]}\sum_{k=1}^{c+1} X_{j,k}^{(c+1)}\right]. \tag{3.151}$$

Now that $c(c+1)\Delta^{(c)}$ has been expressed solely in terms of $\mathbf{X}^{(c+1)}$, we henceforth simplify our notation, writing $\mathbf{X} \equiv \mathbf{X}^{(c+1)}$. Let $Y_j \equiv Y_j^{(c+1)}$ for $j \in [n]$ be the sums of each of the rows of $\mathbf{X}$. Let $\hat{J} \in [n]$ be the random row index $\hat{J} = \hat{\chi}(\mathbf{X})$, where $\hat{\chi}$ is the column invariant row selection rule that selects the smallest index in the (random) set of indices in $[n]$ that maximize Y_j over $j \in [n]$. By this definition:

$$c(c+1)\Delta^{(c)} = \mathbb{E}\left[(c+1)\max_{j\in[n]}\sum_{k=1}^{c} X_{j,k+1} - c\sum_{k=1}^{c+1} X_{\hat{J},k}\right]. \tag{3.152}$$

Further, since the maximum of any set is no smaller than any of its elements, we have the lower bound

$$c(c+1)\Delta^{(c)} \geq \mathbb{E}\left[(c+1)\sum_{k=1}^{c} X_{\hat{J},k+1} - c\sum_{k=1}^{c+1} X_{\hat{J},k}\right]. \tag{3.153}$$

The right hand side may be rearranged as

$$\mathbb{E}\left[\sum_{k=1}^{c} X_{\hat{J},k+1} - cX_{\hat{J},1}\right] = \sum_{k=1}^{c} \mathbb{E}\left[X_{\hat{J},k+1} - X_{\hat{J},1}\right]. \tag{3.154}$$

Suppose the following stochastic ordering condition is true:

$$X_{\hat{J},k+1} >_{\text{st}} X_{\hat{J},1}, \quad k \in [c]. \tag{3.155}$$

This immediately implies $\Delta^{(c)} > 0$, completing the proof.

It remains to prove (3.155). It suffices to show

$$\mathbb{P}(X_{\hat{J},k} > x) > \mathbb{P}(X_{\hat{J},k'} > x) \tag{3.156}$$

for $k, k' \in [c+1]$ with $k > k'$ and $x \in \mathbb{N}$. We start by using the total probability theorem:

$$\begin{aligned}
\mathbb{P}(X_{\hat{J},k} > x) &= \sum_a \sum_j \mathbb{P}(X_{\hat{J},k} > x, X_{\hat{J},k} + X_{\hat{J},k'} = a, \hat{J} = j) \\
&= \sum_a \sum_j \mathbb{P}(X_{j,k} > x, X_{j,k} + X_{j,k'} = a, \hat{J} = j) \\
&= \sum_a \sum_j \mathbb{P}(\hat{J} = j | X_{j,k} > x, X_{j,k} + X_{j,k'} = a) \mathbb{P}(X_{j,k} > x, X_{j,k} + X_{j,k'} = a)
\end{aligned} \tag{3.157}$$

Similarly,

$$\mathbb{P}(X_{\hat{J},k'} > x) = \sum_a \sum_j \mathbb{P}(\hat{J} = j | X_{j,k'} > x, X_{j,k} + X_{j,k'} = a) \mathbb{P}(X_{j,k'} > x, X_{j,k} + X_{j,k'} = a). \tag{3.158}$$

We will establish (3.156) by showing

$$\mathbb{P}(\hat{J} = j | X_{j,k} > x, X_{j,k} + X_{j,k'} = a) \quad = \quad \mathbb{P}(\hat{J} = j | X_{j,k'} > x, X_{j,k} + X_{j,k'} = a) \tag{3.159}$$

$$\mathbb{P}(X_{j,k} > x, X_{j,k} + X_{j,k'} = a) \quad > \quad \mathbb{P}(X_{j,k'} > x, X_{j,k} + X_{j,k'} = a) \tag{3.160}$$

for all x, a, and j. Lemma 7 easily yields (3.159), while Lemma 8 establishes (3.160). □

Lemma 7. *For all $j \in [n]$ and distinct $k, k' \in [c+1]$, the RVs $(\hat{J}, X_{j,k} + X_{j,k'}, X_{j,k}, X_{j,k'})$ form Markov chains*

$$\hat{J} - (X_{j,k} + X_{j,k'}) - X_{j,k}, \quad \hat{J} - (X_{j,k} + X_{j,k'}) - X_{j,k'}. \tag{3.161}$$

Proof. We establish the first chain; the proof of the second is the same. Fix a row index $j \in [n]$, and two distinct column indices $k, k' \in [c+1]$. By the definition of a Markov chain, we must show

$$\mathbb{P}\left(\hat{J} = i \mid X_{j,k} + X_{j,k'} = a, X_{j,k} = x\right) = \mathbb{P}\left(\hat{J} = i \mid X_{j,k} + X_{j,k'} = a\right) \tag{3.162}$$

for all $i \in [n]$ (including $i = j$), integer $a \geq 2$, and integer $x \geq 1$. To make our expressions more compact, we introduce the notation for events $A_1^{(a)} = \{X_{j,k} + X_{j,k'} = a\}$, $A_2 = \{X_{j,k} = x\}$, and $A_{1,2}^{(a)} = A_1^{(a)} \cap A_2$. Then, showing (3.162) is equivalent to showing $\mathbb{P}\left(\hat{J} = i \mid A^{(a)}\right)$ is the same for $A^{(a)}$ equal to either $A_{1,2}^{(a)}$ or $A_1^{(a)}$. Let $\mathbb{N}_{>c} \equiv \{c+1, c+2, \ldots\}$, and define the sets

$$\mathcal{Y}_i = \{\mathbf{y} \in \mathbb{N}_{>c}^n : y_i \geq y_l,\ l \in \{i+1, \ldots, n\},\ y_i > y_l,\ l \in \{1, \ldots, i-1\}\},\ i \in [n]. \tag{3.163}$$

Observe that $\{\hat{J} = i\} = \{\mathbf{Y} \in \mathcal{Y}_i\}$ for each $i \in [n]$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)$ are the row-sums of $\mathbf{X}$. In words, row i is selected under the row selection rule $\hat{\chi}$ as the smallest index among the row-sum maximizing indices precisely when the row-sums lie in (3.163). Here, we use the phrase "row-sum" to indicate the sum of a given row, rather than the sum of the rows. Conditioning on all possible row-sums using the total probability theorem gives:

$$\begin{aligned}
\mathbb{P}(\hat{J} = i | A^{(a)}) &= \sum_{\mathbf{y} \in \mathbb{N}_{>c}^n} \mathbb{P}(\hat{J} = i | \mathbf{Y} = \mathbf{y}, A^{(a)}) \mathbb{P}(\mathbf{Y} = \mathbf{y} | A^{(a)}) \\
&= \sum_{\mathbf{y} \in \mathcal{Y}_i \cap \{\mathbf{y} : y_j \geq a+c-1\}} \mathbb{P}(\hat{J} = i | \mathbf{Y} = \mathbf{y}, A^{(a)}) \mathbb{P}(\mathbf{Y} = \mathbf{y} | A^{(a)}) \\
&= \sum_{\mathbf{y} \in \mathcal{Y}_i \cap \{\mathbf{y} : y_j \geq a+c-1\}} \mathbb{P}(\mathbf{Y} = \mathbf{y} | A^{(a)})
\end{aligned} \tag{3.164}$$

The penultimate equality holds because *i)* $\mathbb{P}(\hat{J} = i | \mathbf{Y} = \mathbf{y}) = 0$ for $\mathbf{y} \notin \mathcal{Y}_i$ and *ii)* $\mathbb{P}(\mathbf{Y} = \mathbf{y} | A^{(a)}) = 0$ for $\mathbf{y}$ with $y_j < a + c - 1$, and the last equality holds because the function $\hat{J}$ of $\mathbf{Y}$ takes value i over

all $\mathbf{y} \in \mathcal{Y}_i$. We now focus on $\mathbb{P}(\mathbf{Y} = \mathbf{y}|A^{(a)})$:

$$\begin{aligned}
\mathbb{P}(\mathbf{Y} = \mathbf{y}|A^{(a)}) &= \mathbb{P}(Y_j = y_j|A^{(a)}) \prod_{l \neq j} \mathbb{P}(Y_l = y_l|A^{(a)}) \\
&= \mathbb{P}(Y_j = y_j|A^{(a)}) \prod_{l \neq j} \mathbb{P}(Y_l = y_l) \\
&= \mathbb{P}\left((X_{j,k} + X_{j,k'}) + \sum_{s \in [c+1] \setminus \{k,k'\}} X_{j,s} = y_j \middle| A^{(a)} \right) \prod_{l \neq j} \mathbb{P}(Y_l = y_l). \quad (3.165)
\end{aligned}$$

The first two equalities hold because of the assumed independence of the rows (note $A^{(a)}$ is specific to row j), and the last equality is by definition of Y_j. Regardless of whether $A^{(a)} = A_1^{(a)}$ or $A^{(a)} = A_{1,2}^{(a)}$, we may write

$$\mathbb{P}(\mathbf{Y} = \mathbf{y}|A^{(a)}) = \mathbb{P}\left(\sum_{s \in [c+1] \setminus \{k,k'\}} X_{j,s} = y_j - a \right) \prod_{l \neq j} \mathbb{P}(Y_l = y_l). \quad (3.166)$$

due to the assumed independence of the columns. Specifically, $(X_{j,s}, s \in [c+1] \setminus \{k, k'\})$ is independent of $(X_{j,k}, X_{j,k'})$. As the above RHS clearly depends upon a but not x, it follows that $\mathbb{P}(\hat{J} = i|A_1^{(a)}) = \mathbb{P}(\hat{J} = i|A_{1,2}^{(a)})$ for each $i \in [n]$, i.e., (3.162) holds. □

Lemma 8. *Consider independent geometric RVs* $X_1 \sim \text{Geo}(q_1)$, $X_2 \sim \text{Geo}(q_2)$, *where* $q_1 \in (0, 1], q_2 \in [0, 1), q_1 > q_2$ *(so* $\mathbb{E}[X_1] < \mathbb{E}[X_2]$*). Then* $\delta(x, a) > 0$ *for all* $x > 0$ *and* $a > x + 1$, *where*

$$\delta(x, a) \equiv \mathbb{P}(X_2 > x, X_1 + X_2 = a) - \mathbb{P}(X_1 > x, X_1 + X_2 = a). \quad (3.167)$$

Proof. Define $\bar{q}_1 = 1 - q_1$ and $\bar{q}_2 = 1 - q_2$. Then:

$$\begin{aligned}
\delta(x,a) &= \sum_{y=x+1}^{a-1} \mathbb{P}(X_2 = y, X_1 = a-y) - \sum_{y=x+1}^{a-1} \mathbb{P}(X_1 = y, X_2 = a-y) \\
&= \sum_{y=x+1}^{a-1} \mathbb{P}(X_2 = y)\mathbb{P}(X_1 = a-y) - \sum_{y=x+1}^{a-1} \mathbb{P}(X_1 = y)\mathbb{P}(X_2 = a-y) \\
&= \sum_{y=x+1}^{a-1} \bar{q}_2^{y-1} q_2 \bar{q}_1^{a-y-1} q_1 - \sum_{y=x+1}^{a-1} \bar{q}_1^{y-1} q_1 \bar{q}_2^{a-y-1} q_2 \\
&= \frac{q_1 q_2 \bar{q}_2^a}{\bar{q}_1 \bar{q}_2} \sum_{y=x+1}^{a-1} \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{a-y} - \frac{q_1 q_2 \bar{q}_2{}^a}{\bar{q}_1 \bar{q}_2} \sum_{y=x+1}^{a-1} \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{y} \\
&= \frac{q_1 q_2 \bar{q}_2^a}{\bar{q}_1 \bar{q}_2} \sum_{y'=1}^{a-x-1} \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{y'} - \frac{q_1 q_2 \bar{q}_2^a}{\bar{q}_1 \bar{q}_2} \sum_{y'=1}^{a-x-1} \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{y'} \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{x} \\
&= \left(1 - \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{x}\right) \frac{q_1 q_2 \bar{q}_2^a}{\bar{q}_1 \bar{q}_2} \sum_{y'=1}^{a-x-1} \left(\frac{\bar{q}_1}{\bar{q}_2}\right)^{y'} > 0. \qquad (3.168)
\end{aligned}$$

□

3.7.5 Proof of Prop. 19

Assume A2: State-independent receptions, heterogeneous receivers. Let the blocklengths c, c' be given, with $c < c'$, and let the workload be $m > c'$. Let $T = T_{n:n}^{(c,m)}$ be the random completion time to transmit all m packets with blocklength c, and let $T' = T_{n:n}^{(c',m)}$ be the time with blocklength c'. Prop. 19 may be restated:

$$T'(\omega) \leq T(\omega),\ \forall \omega \in \Omega \quad \Leftrightarrow \quad c' = kc,\ k \in \{2, 3, 4, \ldots\}. \qquad (3.169)$$

We first prove the forward direction (necessity): $T'(\omega) \leq T(\omega),\ \forall \omega \in \Omega \Rightarrow c' = kc$ for integer $k \geq 2$, which is equivalent to: if $c' = kc + l$ with integer $k \geq 1$ and integer $l \in \{1, \ldots, c-1\}$ then there exists $\omega : T'(\omega) > T(\omega)$. We begin with perhaps the simplest example showing that increasing the blocklength may increase the completion time for some sample paths. In particular, consider $n = 2$ receivers and a workload of $m = 4$ packets which are transmitted using one of two schemes: *i*) a blocklength of $c = 2$ for a total of 2 full blocks, and *ii*) a blocklength of $c' = 3$ for a total of 1 full and 1 partial block, as shown in Fig. 3.5. The realization ω for the erasure channels is illustrated at

the top of the figure for the first 6 time slots, where s (f) indicates a successful (failed) transmission for that receiver in that time slot. For this realization, the longer blocklength construction finishes after the shorter blocklength construction. The times t_a, t_b, t_c, t_d, t_e are defined in the subsequent generalization of this example.

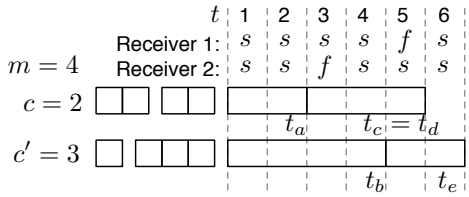

Figure 3.5: Simple example illustrating that increasing the blocklength may increase the completion time for some sample paths.

We now generalize the above example; see Fig. 3.6. Let there be $n = 2$ receivers. By assumption, the larger blocklength is $c' = kc + l$ for integer $k \geq 1$ and integer $l \in \{1, \ldots, c-1\}$. The blocks under both blocking constructions are shown as rectangles in the figure, and the numbers inside the blocks are the block indices. Give each receiver a counter initialized at zero, and increment each receiver's counter upon a *useful* reception at that receiver, where a reception is useful if that receiver has not yet received a blocklength of receptions within the current block. These quantities are shown for the two receivers under the two blocking constructions for the various intervals of time shown in the figure. The realization ω of the erasure channels for each receiver is shown at the top of the figure, with s (f) indicating success (failure), respectively. Under this realization, the completion time of block index k under blocklength c is time $t_a = kc$. As of time t_a, the first block under blocklength c' has received kc out of the $kc + l$ receptions needed to complete the block. Let time $t_b = t_a + c$ be the time of completion of the first block under blocklength c'. Next, let time $t_c = t_b + c - l$ be the time of completion of block $k + 1$ under blocklength c. Let $t_d = m + c - l$ be the completion time of the workload m under blocklength c, and $t_e = t_d + c - l$ be the workload completion time under blocklength c'. The key insight is this: the non-overlapping $c - l$ failed receptions for each of the two receivers occur in the same block (block $k + 1$) under blocklength c (and therefore the duration of the block is delayed by $c - l$), while under blocklength c' the failed receptions for the two receivers occur in different blocks, which delays *both* of its first two blocks by $c - l$.

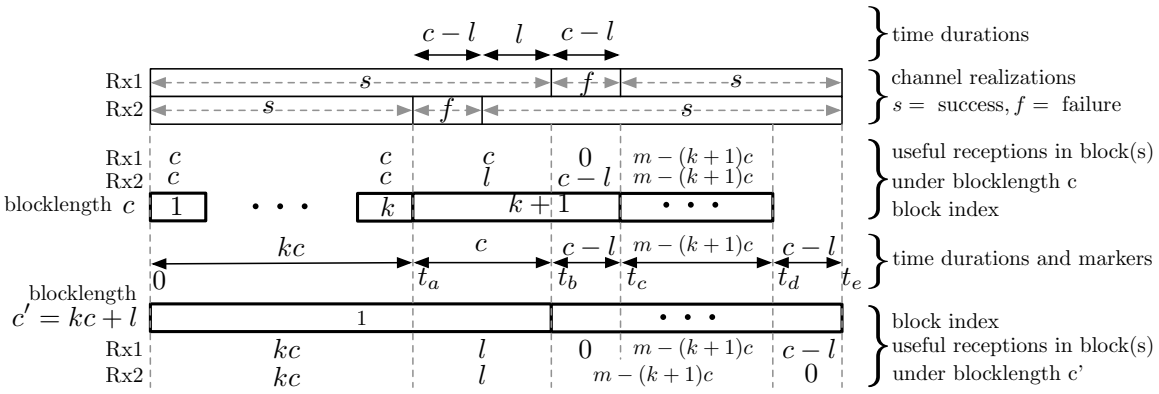

Figure 3.6: For any blocklengths c, c' with $c' > c$ and $c' = kc + l$ with $l \in \{1, \ldots, c-1\}$ and workload $m > c'$, there exists a realization under which the longer blocklength construction will have a longer completion time.

We next prove the reverse direction (sufficiency): if $c' = kc$ for some integer $k \geq 2$, then $T'(\omega) \leq T(\omega)\ \forall \omega \in \Omega$. See Fig. 3.7. This proof is essentially an extension of the proof of Lemma 1. Fix the realization ω; in what follows we suppress the dependence upon ω. By assumption $c' = kc$ for some integer $k \geq 2$ and $m > c'$. Write $m = k'c' + l'$ for integer $k' \geq 1$ and integer $l' \in \{0, \ldots, c'-1\}$, so that the workload m requires $k'+1$ blocks under blocklength c': k full blocks of size c' and, if $l' > 0$, a partial block of size $l' < c'$. Let $(y^{(c')}_{n:n,1}, \ldots, y^{(c')}_{n:n,k'}, y^{(l')}_{n:n,k'+1})$ denote the durations of time required to complete the $k'+1$ blocks under blocklength c', and $(t^{(c')}_{n:n,1}, \ldots, t^{(c')}_{n:n,k'}, t^{(c')}_{n:n,k'+1})$ the corresponding sequence of partial sums, so that $t^{(c')}_{n:n,\iota}$ is the completion time of block ι under blocklength c', and $t^{(c')}_{n:n,k'+1}$ the completion time of the workload. Write $l' = k^{''}c + l$ for integer $k^{''} \geq 0$ and integer $l \in \{0, \ldots, c-1\}$, so that workload m requires $k'k+k^{''}+1$ blocks under blocklength c: $k'k+k^{''}$ blocks of size c and, if $l > 0$, a partial block of size $l < c$. Similarly, let $(y^{(c)}_{n:n,1}, \ldots, y^{(c)}_{n:n,k'k+k^{''}}, y^{(l)}_{n:n,k'k+k^{''}+1})$ denote the durations of time required to complete the $k'k + k^{''} + 1$ blocks under blocklength c, and $(t^{(c)}_{n:n,1}, \ldots, t^{(c)}_{n:n,k'k+k^{''}}, t^{(c)}_{n:n,k'k+k^{''}+1})$ the corresponding sequence of partial sums, so that $t^{(c)}_{n:n,\iota}$ is the completion time of block ι under blocklength c, and $t^{(c)}_{n:n,k'k+k^{''}+1}$ the completion time of the workload. We must show $t^{(c')}_{n:n,k'+1} \leq t^{(c)}_{n:n,k'k+k^{''}+1}$.

We first show $t^{(c')}_{n:n,\iota} \leq t^{(c)}_{n:n,k\iota}$ for each $\iota \in \{1, \ldots, k'\}$ by induction. When $\iota = 1$, $t^{(c')}_{n:n,1} \leq t^{(c)}_{n:n,k}$ can be verified by using the same ideas used in proving Lem. 1. Next, assuming $t^{(c')}_{n:n,\iota-1} \leq t^{(c)}_{n:n,k(\iota-1)}$ and denoting $t^{(c')}_{j,\iota} = \min\{t : r_j(t^{(c')}_{n:n,\iota-1}, t] = c' = kc\}$ where $r_j(t_1, t2]$ is the number of successful

(not necessarily useful) receptions by receiver j during the time interval $(t_1, t_2]$, we have:

$$t_{j,\iota}^{(c')} \overset{(a)}{\leq} \min\{t : r_j(t_{n:n,k(\iota-1)}^{(c)}, t] = c' = kc\} \overset{(b)}{\leq} t_{n:n,k\iota}^{(c)}, \tag{3.170}$$

where (a) is due to the induction hypothesis and (b) is seen to be true by observing $r_j(t_{n:n,k(\iota-1)}^{(c)}, t_{n:n,k\iota}^{(c)}] \geq kc$ for all $j \in [n]$. So $t_{n:n,\iota}^{(c')} = \max_j t_{j,\iota}^{(c')} \leq t_{n:n,k\iota}^{(c)}$, finishing the induction step. Finally, for the last partial block under c', a similar argument (again to the one used in proving Lem. 1, as effectively the remaining workload does not exceed the larger blocklength c') together with the just proved result $t_{n:n,\iota}^{(c')} \leq t_{n:n,k\iota}^{(c)}$ (specialized with $\iota = k'$) give $t_{n:n,k'+1}^{(c')} \leq t_{n:n,k'k+k''+1}^{(c)}$.

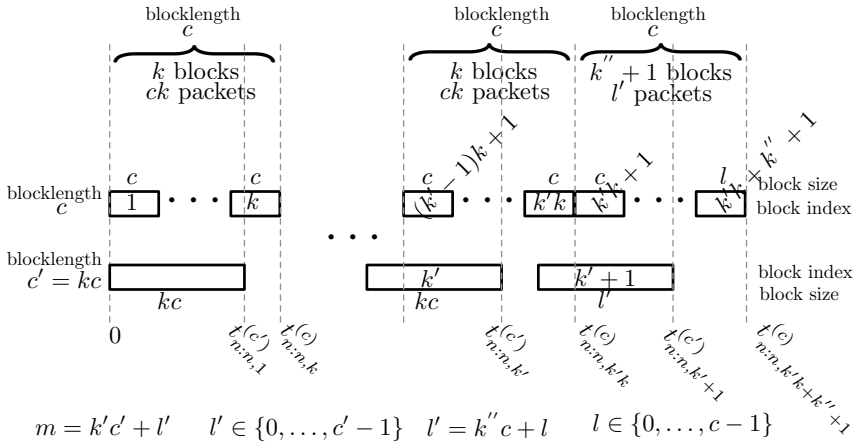

Figure 3.7: For any blocklengths c, c' with $c' > c$ and $c' = kc$ and workload $m > c'$, the longer blocklength construction will have a completion time no longer than the shorter blocklength construction.

3.7.6 Lower and upper bounds on the expected maximum order statistic

This appendix presents two inequalities on the expected maximum order statistic, $\mathbb{E}[Z_{n:n}]$: a lower bound due to de la Cal and Cárcamo[3] and an upper bound due to Ross and Peköz[2]. We present the inequalities, then give a pertinent example of their application, including an illustration of how to use the bounds to establish asymptotic tightness.

Proposition 26 ([3] Thm. 13)**.** *Let $(Z_1, \ldots, Z_n)$ be independent and identically distributed (not necessarily nonnegative) RVs. Then:*

$$\mathbb{E}[Z_{n:n}] \geq t - (t - \mu_Z) F_Z^{n-1}(t). \tag{3.171}$$

for all $t \geq \mu_Z$, where μ_Z, F_Z denote the mean and CDF of each Z_j.

Prop. 26 can be extended to the case of RVs that are independent but not necessarily identically distributed.

Proposition 27 ([2] §4.6)**.** *Let $(Z_1, \ldots, Z_n)$ be nonnegative (not necessarily independent) RVs. Then:*

$$\mathbb{E}[Z_{n:n}] \leq s + \sum_{j=1}^{n} \int_s^\infty (1 - F_{Z_j}(z)) \mathrm{d}z, \tag{3.172}$$

for all $s \geq 0$, where F_{Z_j} is the CDF of Z_j. The bound is tightest for s^* satisfying

$$\sum_{j=1}^{n} (1 - F_{Z_j}(s^*)) = 1. \tag{3.173}$$

Since the left side of (3.173) is a decreasing function in s, it follows that i) the solution of (3.173) is unique, and ii) may be easily found numerically via bisection search. Both bounds can be (or have been) shown to have a counterpart for bounding the expected minimum order statistic $\mathbb{E}[Z_{1:n}]$. Observe that both inequalities have a degree of freedom in the form of a parameter that may be chosen either to make the bound as tight as possible, or put the bound in a certain form. We illustrate the bounds for the exponential case. In particular, let $(\tilde{X}_1, \ldots, \tilde{X}_n)$ be iid exponential RVs with unit rate. There is no loss in generality in setting the rate $\lambda = 1$ since $\frac{1}{\lambda}\tilde{X} \sim \mathrm{Exp}(\lambda)$, and in particular $\mathbb{E}[\max_{j \in [n]} \mathrm{Exp}_j(\lambda)] = \frac{1}{\lambda}\mathbb{E}[\tilde{X}_{n:n}]$.

Proposition 28. *Let $(\tilde{X}_1, \ldots, \tilde{X}_n)$ be iid unit rate exponential RVs. Then the functions $l_{\tilde{X}}(t, n)$ and $u_{\tilde{X}}(s, n)$ below satisfy $l_{\tilde{X}}(t, n) \leq \mathbb{E}[\tilde{X}_{n:n}] \leq u_{\tilde{X}}(s, n)$ for all $s \geq 0$, $t \geq 1$, and $n \in \mathbb{N}$, where:*

$$\begin{aligned} l_{\tilde{X}}(t, n) &\equiv t - (t-1)(1 - \mathrm{e}^{-t})^{n-1} \\ u_{\tilde{X}}(s, n) &\equiv s + n\mathrm{e}^{-s} \end{aligned} \tag{3.174}$$

For the choice $t_n = 1 + \log n - \varepsilon_n$ where $\varepsilon_n = o(\log n)$ and $\varepsilon_n = \omega(1)$ we have

$$l_{\tilde{X}}(t_n, n) \equiv l_{\tilde{X}}(n) = 1 + (\log n - \varepsilon_n) \left(1 - \left(1 - \frac{1}{n\mathrm{e}} \mathrm{e}^{\varepsilon_n}\right)^{n-1}\right). \tag{3.175}$$

Furthermore, the optimal value of s is $s_n^* = \log n$, for which

$$u_{\tilde{X}}(s_n^*, n) \equiv u_{\tilde{X}}(n) = 1 + \log n. \tag{3.176}$$

Finally, the bounds $(l_{\tilde{X}}(n), u_{\tilde{X}}(n))$ are asymptotically tight in n in that, as $n \to \infty$,

$$\frac{l_{\tilde{X}}(n)}{\log n} \to 1, \quad \frac{u_{\tilde{X}}(n)}{\log n} \to 1. \tag{3.177}$$

These together imply $\mathbb{E}[\tilde{X}_{n:n}]/\log n \to 1$.

Proof. Specializing the lower (upper) bound in Prop. 26 (27), respectively, to the exponential case gives (3.174). Substituting the given form for t_n into the lower bound gives (3.175). Differentiation of the convex function $u_{\tilde{X}}(s, n)$ with respect to s, equating with zero, and solving for s gives $s_n^* = \log n$, and $u_{\tilde{X}}(n)$ in (3.176). The limit of $u_{\tilde{X}}(n)/\log n$ as $n \to \infty$ is immediate. It remains to establish that the limit of

$$\frac{l_{\tilde{X}}(n)}{\log n} = \frac{1 + (\log n - \varepsilon_n)\left(1 - \left(1 - \frac{1}{n\mathrm{e}}\mathrm{e}^{\varepsilon_n}\right)^{n-1}\right)}{\log n} = \frac{1}{\log n} + \left(1 - \frac{\varepsilon_n}{\log n}\right)\left(1 - \left(1 - \frac{1}{n\mathrm{e}}\mathrm{e}^{\varepsilon_n}\right)^{n-1}\right) \tag{3.178}$$

as $n \to \infty$ is 1. Since $\varepsilon_n = o(\log n)$, it suffices to show that the limit of

$$\left(1 - \frac{1}{n\mathrm{e}}\mathrm{e}^{\varepsilon_n}\right)^{n-1} = \left(1 - \frac{\mathrm{e}^{\varepsilon_n - 1}}{n}\right)^{n-1} = \left[\left(1 - \frac{\mathrm{e}^{\varepsilon_n - 1}}{n}\right)^{n\mathrm{e}^{1-\varepsilon_n}}\right]^{\frac{n-1}{n}\mathrm{e}^{\varepsilon_n - 1}} \tag{3.179}$$

as $n \to \infty$ is zero. First consider $\mathrm{e}^{\varepsilon_n - 1}/n$. For any positive sequence $\{a_n\}$:

$$\lim_{n\to\infty} \frac{\log a_n}{\log n} = -1 \;\Rightarrow\; \lim_{n\to\infty} \log a_n = -\infty \;\Leftrightarrow\; \lim_{n\to\infty} a_n = 0. \tag{3.180}$$

For $a_n = \mathrm{e}^{\varepsilon_n - 1}/n$ with $\varepsilon_n = o(\log n)$ this gives directly that $\mathrm{e}^{\varepsilon_n - 1}/n \to 0$. Recall $\lim_{b\to\infty}(1 - 1/b)^b =$

e^{-1}. Choosing $b = n\mathrm{e}^{1-\varepsilon_n}$ it follows that

$$\lim_{n\to\infty}\left(1-\frac{\mathrm{e}^{\varepsilon_n-1}}{n}\right)^{n\mathrm{e}^{1-\varepsilon_n}} = \mathrm{e}^{-1}. \tag{3.181}$$

Finally, the limit as $n\to\infty$ of the logarithm of the rightmost expression in (3.179) equals (applying (3.181)):

$$\left(\lim_{n\to\infty}\frac{n-1}{n}\mathrm{e}^{\varepsilon_n-1}\right)\left(\lim_{n\to\infty}\log\left(1-\frac{\mathrm{e}^{\varepsilon_n-1}}{n}\right)^{n\mathrm{e}^{1-\varepsilon_n}}\right) = -\lim_{n\to\infty}\frac{n-1}{n}\mathrm{e}^{\varepsilon_n-1} = -\infty. \tag{3.182}$$

It follows that the limit of (3.179) as $n\to\infty$ is zero.

□

Using the min-max identity (Prop. 2), the fact that $\min_{j\in\mathcal{A}}\tilde{X}_j \sim \mathrm{Exp}(s)$ for any $\mathcal{A}\in[n]_s$, and the binomial coefficient absorbtion identity gives

$$\mathbb{E}[\tilde{X}_{n:n}] = \sum_{s=1}^{n}(-1)^{s+1}\sum_{\mathcal{A}\in[n]_s}\mathbb{E}\left[\min_{j\in\mathcal{A}}\tilde{X}_j\right] = \sum_{s=1}^{n}(-1)^{s+1}\binom{n}{s}\frac{1}{s} = \sum_{s=1}^{n}\frac{1}{s} = H_n, \tag{3.183}$$

where H_n denotes the n^{th} harmonic number. We select $\varepsilon_n = \log\log n$ under which the lower bound becomes

$$l_{\tilde{X}}(n) = 1 + (\log n - \log\log n)\left(1-\left(1-\frac{\log n}{n}\right)^{n-1}\right). \tag{3.184}$$

These bounds are shown in Fig. 3.8.

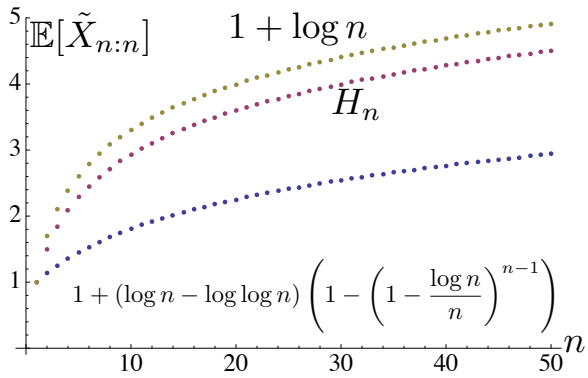

Figure 3.8: The lower bound (blue), upper bound (yellow), and exact value (red) for $\mathbb{E}[\tilde{X}_{n:n}]$, for $\tilde{X}_1, \ldots, \tilde{X}_n$ iid unit rate exponentials.

Chapter 4: Conclusion

This thesis has addressed both the uplink (§2) and the downlink (§3) of (one-hop) wireless networks. Specifically, in Chapter 2 we have established new properties of the set Λ, an inner bound on the Aloha stability region, including membership testing, equivalent forms, stabilizing controls and bounds, with a focus on establishing inner and outer bounds induced by simple polyhedra, spheres, and ellipsoids. These non-parametric bounds partially inform the structure of the set Λ, and might be sufficient for testing membership in Λ for a given arrival rate vector $\mathbf{x}$. Exact membership testing can be performed by checking the existence of a positive root of a certain polynomial equation (Prop. 1), or by solving a convex program (Cor. 3) which also finds the set of stabilizing controls. The critical stabilizing control(s) can be obtained by using the augmented root testing (Prop. 3).

Natural extensions of this chapter include the following. First, to improve the augmented root testing (Prop. 3) such that both of the two positive roots (or two critical stabilizing controls) can be expressed explicitly. Currently, given a critical stabilizing control $\mathbf{p}$, only one of the two positive roots is explicitly shown to equal $1/\pi(\mathbf{p})$. To obtain the other positive root (critical stabilizing control) we need to resort to a procedure demonstrated in diagram (2.166). Second, to identify classes of $\mathcal{E}(c, a_1, a_2)$ ellipsoids different from the families we have constructed that induce tighter bounds on Λ. Third, to seek a closed-form expression for the volume of Λ that is simpler and easier to compute than the one we have provided, and/or to provide a good approximation of how the volume varies with n.

In Chapter 3 we have investigated the random block delay, $Y_{n:n}^{(c)}$, when broadcasting c packets over independent erasure channels to n receivers with reception probabilities $\mathbf{q}$ using both uncoded transmission and random linear coding. Key results here involve bounds, exact expressions, recurrences, and asymptotic properties (in c, or in n) for the r^{th} ($r \in \mathbb{N}$) moment of $Y_{n:n}^{(c)}$. A fundamental insight is seen from Props. 16 and 17, suggesting the tradeoff between the expected per packet delay and the decoding delay: on one hand, increasing blocklength always reduces the expected per packet

delay (or total delay if the workload is fixed, recall Cor. 4); on the other hand, since the decoding is performed on a block-by-block basis, this implies information packets with earlier timestamps risk being expired if the blocklength c is too large.

Several extensions of this chapter seem natural. First and foremost, our results in §3.6 study the delay as n grows large for fixed c, while the results in[37] have established that $c = c(n)$ should grow with n according to $\Omega(\log n)$ in order to have a non-dimensioning throughput (a phase transition occurs at $c(n) = \Theta(\log n)$). A straightforward suggestion is to use our framework of order statistic inequalities, stochastic ordering, and extreme value theory to address this case. Second, there are additional results that are desirable in §3.5. In particular, *i*) in §3.5.2 we conjecture the expected per packet delay is not only decreasing in c, but in fact convex decreasing in c, and *ii*) in §3.5.3 we believe a better characterization of a "good" choice of t_c in the lower bound on $\mathbb{E}\left[Y_{n:n}^{(c)}\right]/c$ should be possible.

Bibliography

[1] A. Eryilmaz, A. Ozdaglar, M. Medard, and E. Ahmed. On the delay and throughput gains of coding in unreliable networks. *IEEE Transactions on Information Theory*, 54(12):5511–5524, December 2008.

[2] S. M. Ross and E. A. Pekoz. *A Second Course in Probability*. Probability Bookstore, 2007.

[3] J. de la Cal and J. Cárcamo. Inequalities for expected extreme order statistics. *Statistics and Probability Letters*, 73:219–231, 2005.

[4] Norman Abramson. The ALOHA system: another alternative for computer communications. In *Proceedings of the fall joint computer conference*, pages 281–285. ACM, 1970.

[5] R.R. Rao and A. Ephremides. On the stability of interacting queues in a multiple–access system. *IEEE Transactions on Information Theory*, 34(5):918–930, September 1988.

[6] W. Luo and A. Ephremides. Stability of n interacting queues in random–access systems. *IEEE Transactions on Information Theory*, 45(5):1579–1587, July 1999.

[7] S.C. Kompalli and R.R. Mazumdar. On the stability of finite queue slotted-Aloha protocol. *IEEE Transactions on Information Theory*, 59(10):6357–6366, October 2013.

[8] Lawrence G. Roberts. ALOHA packet system with and without slots and capture. *ACM SIGCOMM Computer Communication Review*, 5(2):28–42, April 1975.

[9] N. Abramson. The throughput of packet broadcasting channels. *IEEE Transactions on Communications*, 25(1):117–128, January 1977.

[10] B.S. Tsybakov and V.A. Mikhailov. Ergodicity of the slotted Aloha system. *Problemy Peredachi Informatsii*, 15(4):73–87, 1979.

[11] W. Szpankowski. Stability conditions for some distributed systems: buffered random access systems. *Advances in Applied Probability*, 26(2):498–515, June 1994.

[12] V. Anantharam. The stability region of the finite-user slotted Aloha protocol. *IEEE Transactions on Information Theory*, 37(3):535–540, May 1991.

[13] Charles Bordenave, David McDonald, and Alexandre Proutiere. Performance of random medium access control, an asymptotic approach. In *SIGMETRICS '08: Proceedings of the 2008 ACM SIGMETRICS international conference on Measurement and modeling of computer systems*, pages 1–12, New York, NY, USA, 2008. ACM. ISBN 978-1-60558-005-0. doi: http://doi.acm.org/10.1145/1375457.1375459.

[14] J. Luo and A. Ephremides. On the throughput, capacity and stability regions of random multiple access. *IEEE Transactions on Information Theory*, 52(6):2593–2607, June 2006.

[15] Vijay G. Subramanian and Douglas J. Leith. On the rate region of CSMA/CA WLANs. *IEEE Transactions on Information Theory*, 59(6):3932–3938, June 2013.

[16] J.L. Massey and P. Mathys. The collision channel without feedback. *IEEE Transactions on Information Theory*, 31(2):192–204, March 1985.

[17] K.A. Post. Convexity of the nonachievable rate region for the collision channel without feedback. *IEEE Transactions on Information Theory*, 31(2):205–206, March 1985.

[18] Dimitri Bertsekas and John Tsitsiklis. *Introduction to Probability*. Athena Scientific Press, 2nd edition, 2008.

[19] A. Grundmann and H.M. Moller. Invariant integration formulas for the n-simplex by combinatorial methods. *SIAM Journal on Numerical Analysis*, 15(2):282–290, April 1978.

[20] H. Wilf. *Generatingfunctionology*. A.K. Peters, Wellesley, MA, 3rd edition, 1994.

[21] Richard S. Ellis. Volume of an n-simplex by multiple integration. *Elemente Der Mathematik*, 31(3):57–59, 1976.

[22] Necdet Batir. Sharp inequalities for factorial n. *Proyecciones Journal of Mathematics*, 27(1): 97–102, May 2008.

[23] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[24] Osman Güler and Filiz Gürtuna. The extremal volume ellipsoids of convex bodies, their symmetry properties, and their determination in some special cases. *arXiv preprint arXiv:10709.0707v1*, 2007.

[25] D. Mortari. On the rigid rotation concept in n-dimensional spaces. *Journal of the Astronautical Sciences*, 49(3), July-September 2001.

[26] Dan Sloughter. The calculus of functions of several variables. URL `http://www.synechism.org/wp/the-calculus-of-functions-of-several-variables/`. (accessed on Aug. 14, 2014).

[27] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer, second edition, 2011.

[28] W. Erwin Diewert, Mordechai Avriel, and Israel Zang. Nine kinds of quasiconcavity and concavity. *Journal of Economic Theory*, 25(3):397–420, December 1981.

[29] Alberto Cambini and Laura Martein. *Generalized Convexity and Optimization: Theory and Applications*, volume 616 of *Lecture notes in economics and mathematical systems*. Springer, 2009.

[30] Shashi Kant Mishra and Giorgio Giorgi. *Invexity and Optimization*, volume 88 of *Nonconvex optimization and its applications*. Springer, 2008.

[31] T. Ho, M. Médard, R. Koetter, D.R. Karger, M. Effros, J. Shi, and B. Leong. A random linear network coding approach to multicast. *IEEE Transactions on Information Theory*, 52 (10):4413–4430, October 2006.

[32] J.W. Byers, M. Luby, and M. Mitzenmacher. A digital fountain approach to asynchronous reliable multicast. *IEEE Journal on Selected Areas in Communications (JSAC)*, 20(8):1528–1540, August 2002.

[33] R. Cogill, B. Shrader, and A. Ephremides. Stable throughput for multicast with random linear coding. *IEEE Transactions on Information Theory*, 57(1):267–281, January 2011.

[34] R. Cogill and B. Shrader. Multicast queueing delay: Performance limits and order-optimality of random linear coding. *IEEE Journal on Selected Areas in Communications*, 29(5):1075–1083, May 2011.

[35] R. Cogill and B. Shrader. Delay bounds for random linear coding in multihop relay networks. In *Proceedings of the 2011 Conference on Information Science and Systems (CISS)*, 2011.

[36] Y. Yang and N.B. Shroff. Throughput of rateless codes over broadcast erasure channels. In *ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 125–133, Hilton Head Island, SC, June 2012.

[37] B.T. Swapna, A. Eryilmaz, and N.B. Shroff. Throughput-delay analysis of random linear network coding for wireless broadcasting. *IEEE Transactions on Information Theory*, 59(10): 6328–6341, October 2013.

[38] A. Eryilmaz, A. Ozdaglar, and M. Médard. On delay performance gains from network coding. In *40th Annual Conference on Information Sciences and Systems (CISS)*, pages 864–870, Princeton, NJ, USA, March 2006.

[39] Sheldon M. Ross. *Stochastic Processes*. John Wiley & Sons, 2nd edition, 1996.

[40] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[41] Robert Milson. Stirling numbers of the second kind, March 2013. URL `http://planetmath.org/sites/default/files/texpdf/32805.pdf`. (accessed on Aug. 14, 2014).

[42] V. M. Zaskulnikov. Statistical mechanics of fluids in a step potential. *arXiv preprint arXiv:1205.6546v1*, 2012.

[43] Bennett Eisenberg. On the expectation of the maximum of iid geometric random variables. *Statistics & Probability Letters*, (78):135–143, 2008.

[44] W. Szpankowski and V. Rego. Yet another application of binomial recurrence: order statistics. *Computing*, (43):401–410, 1990.

[45] T. K. Dikaliotis, A G. Dimakis, T. Ho, and M. Effros. On the delay advantage of coding in packet erasure networks. In *Proceedings of the IEEE Information Theory Workshop (ITW)*, 2010.

[46] P. J. Grabner and H. Prodinger. Maximum statistics of n random variables distributed by the negative binomial distribution. *Combinatorics, Probability and Computing*, 6(2):179–183, 1997.

[47] Allan Gut. *An Intermediate Course in Probability*. Springer, 2nd edition, 2009.

[48] S. Resnick. *Extreme Values, Regular Variations, and Point Processes*. Springer-Verlag, New York, 1987.

[49] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York, 1983.

[50] H.A. David and H.N. Nagaraja. *Order Statistics*. Wiley, third edition, 2003.

Vita

Nan Xie (谢南) received his B.S. (2003) in communications engineering and M.S. (2006) in circuits and systems, both from Wuhan University, China. He attended the University of Florida from 2006 to 2007 where he was also a graduate assistant. Starting in the Fall of 2008, he has been working towards a Ph.D. in electrical engineering at the Electrical and Computer Engineering Department of Drexel University. In the fields of communications and networking, he is particularly interested in a fundamental understanding of theoretical problems whose solutions could shed light on design and performance optimization. He may be reached at nan.xie@drexel.edu or nonshell@gmail.com.